

Basic research: an economic good

A new survey of the literature of economics and social aspects of science gives support to commonly held beliefs about the value of 'blue-skies' research.

As sciences go, economics may be dismal, while in many people's eyes the sociology of science can seem downright absurd. But when both are used to show that government support for basic science is economically an excellent thing, respect for them can only soar. How testable is that conclusion — and how strong?

The Science Policy Research Unit at the University of Sussex has performed a useful service by surveying the literature on the benefits of government support for basic research, and on related policies in various countries. (The report — *The Relationship between Publicly Funded Basic Research and Economic Performance*, by Ben Martin *et al.* — was sponsored by the UK Treasury.) The report describes the state of the evidence by which economic benefits can be tested. The result, encouragingly, is support for ideas that most researchers, but not everybody in finance ministries, would have wishfully taken for granted. When examined critically, they emerge all the stronger.

The idea that the knowledge that emerges from basic research is its principal — let alone its sole — public benefit is becoming more questionable. Just how public is the benefit? That knowledge, once published, can be picked up by others only partially, and then not for free, given the trained individuals and specialized equipment required. Scientists may take this for granted, but economists have underestimated the costs of this 'public' knowledge by concentrating on costs of production and neglecting those of use. The role of information technology is ambiguous: the Internet is enhancing the availability of information, but increased use of computation and simulation strengthens the less publicly available but no less economically significant product of basic research, the tacit skills of researchers. And the evidence is there to highlight the economic usefulness of ex-researchers in esoteric science as much as those in catalysis and biotechnology.

What emerge clearly from this survey are the persistent uncertainties in attempts to quantify the economic returns on public expenditure on basic research — the benefits arising from the use of the knowledge gained and in the employment of those trained. Work by Edwin Mansfield, who examined the industrial utilization of academic research in 75 major US companies, is much quoted — and naively so, to judge by the report's list of "heroic" assumptions made in Mansfield's study. But the firm conclusion remains that the economic returns on society's investment are substantiated and substantial, and that several per cent of the companies' sales would not have been achieved without university research.

Other benefits can be analysed. Surveys of industrialists clarify the importance of the transfer of instrumentation, which can underpin new disciplines or can amount to an industry in itself. But the economic benefits remain to be quantified. More progress has been made on the impact of basic research on the ability of business to find new solutions to technological problems. In biotechnology, the impact of basic research is direct. In other disciplines, the path from discovery to innovation is less clear cut. But the evidence points to substantial direct and indirect benefits nevertheless. And a clear advantage of geographical proximity in industrial/academic collaboration has emerged. A country that wishes to stimulate high-tech industrial growth would do particularly well to encourage strong regional networks.

To get the best out of such networks, companies must support

research themselves, have their employees publish original work, and be in contact with the community. And the idea is emerging that the research network itself should be seen as a significant unit of economic public good.

Little of this will surprise those whose occupation is research. But providing tangible evidence lessens one's dependence on faith. The unsurprising disappointment in the report, acknowledged by the authors themselves, is that, while giving nothing but comfort to the idea that basic research is economically a good thing, it gives no help at all to governments trying to assess the benefits of one level of funding rather than another. □

In search of anger

Road transport should be on trial. New proposals on air quality suggest that it is not.

CARBON monoxide, nitrogen dioxide, sulphur dioxide, benzene, 1,3 butadiene, lead and poison-coated soot. As a result of some of those, ozone. And hence bronchial illness, lymphomas, leukaemia, heart attacks....

There is a much that road transport has to answer for. But however proudly it might proclaim its introduction of the tightest air-quality standards in Europe (come 2005), the United Kingdom government, like so many others, has manifestly failed to bring the manufacturers and owners of cars, trucks and buses to book. Its "air quality strategy" announced last week (see page 743) does not deserve that grandiloquent title. It is dependent on future regulation by the European Union, and undermined by a stated intention to take into account the costs of any proposed initiatives: it will invite manufacturers and haulage associations to spell out such costs in detail before decisions are taken on how the new standards — some of which, it turns out, are droppable — will be achieved. More determination is required than that.

The first priority must be the policing of emissions from older vehicles. Any car three years old or more should have to pass an annual test of its exhaust, while traffic police should have the power to stop and inspect vehicles for dirty emissions and, if necessary, issue fume tickets with fines proportional to the scale of excess.

Another priority is the imposition of additional cost on car use. Here the UK government has done something tangible, with steadily increasing duties on petrol. But the continuing use of company cars as a perk should cease, while financial penalties should be suffered by those who drive in cities out of convenience rather than necessity. Manufacturers and purchasers of cars will have to accept an increase in costs too.

Such initiatives would be perceived to make sense by a public more conscious of what it is suffering. People ingesting soot, sulphur dioxide and the rest of it should be angrier than they are. But the only time Europeans, at least, have got angry about petrol was when an oil rig was about to be allowed to sink to the ocean floor. If that is a test of the achievements of green organizations, then they have failed indeed. □



British government releases plans to cut air pollution by 2005

London. The UK government has unveiled its National Air Quality Strategy and promised that by 2005 British cities will be rid of smog and dangerous levels of air pollutants. The strategy is seen as a key attempt by the government to set out its 'green'

Local authorities will be given new powers to pedestrianize streets. They may also be allowed to penalize drivers with suspect emissions as well as those found parking in pollution hot-spots. Industry, meanwhile, will be encouraged to develop new emissions technologies, such as catalytic converters for diesel engines.

John Gummer, secretary of state for the environment, says that the responsibility for reducing pollution is a "partnership" in which the public also has a stake. Gummer says he envisages families owning fewer cars and using them less often.

The document has had a mixed reception. The oil and motoring lobbies are relieved at the decision not to levy more taxes, and to encourage new technologies instead. The decision to defer measures to curb diesel engine exhaust fumes, which account for 80 per cent of particulate matter emissions from road traffic, will also be welcomed by the road haulage lobby.

But Sir John Houghton, chairman of the Royal Commission on Environmental Pollution, believes the government should not delay measures to cut diesel emissions. A 1994 report from the Commission on transport and the environment suggested incentives be given to owners of old diesels to buy newer and cleaner ones. "This is an important issue, which the government does not

seem to have picked up," says Houghton.

The strategy acknowledges that emissions of particulates and nitrogen oxides can be cut by up to 70 per cent for each diesel vehicle that is converted to run on alternative fuels such as compressed natural gas or liquefied petroleum gas. Roy Harrison, professor of environmental health at the University of Birmingham, and a member of EPAQS, says such changes ought to be encouraged, and will require minimal modifications to engines.

Environmental groups and opposition political parties have dismissed the strategy as 'too little, too late'. The Labour party says it is an acknowledgement that previous policies — under Margaret Thatcher — to invest in the building of roads as opposed to public transport, were a mistake. But Labour has yet to unveil its own proposals, which are unlikely to include measures such as taxation, because of the opposition party's determination not to announce new taxes in the run-up to a general election.

Tony Bosworth, air pollution campaigner for Friends of the Earth, says the government strategy does not address the main cause of air pollution: "too much traffic". The numbers of cars, he says, are projected to double by 2025. Reducing this rate of growth, he believes, should be a higher priority.

Ehsan Masood

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Bill Stephenson/Panos

Yorkshire smog cloaks Sheffield.

credentials, ahead of the opposition Labour party, which has yet to formulate its own proposals.

The strategy document sets out to reduce atmospheric concentrations of eight classes of pollutants that are considered partly to blame for rising cases of asthma and premature deaths during smog episodes. The targets were recommended over the past two years by the government's Expert Panel on Air Quality Standards (EPAQS).

The strategy, however, does not spell out how the objectives will be met. For example, it does not say how the government will reduce concentrations of the tiny soot 'particulates' from vehicle exhaust emissions that are linked to heart attacks, and breathing difficulties. The UK Department of Health estimates that particulates are responsible for several thousand premature deaths each year and 10–20,000 hospital admissions.

The government believes its current policies will meet targets to reduce concentrations of four of the pollutants: carbon monoxide, benzene, 1,3 butadiene and lead. However, it acknowledges that "additional measures" — yet to be announced — will be needed to meet targets for the remaining four: nitrogen dioxide, particulates, ozone and sulphur dioxide.

The air quality strategy also indicates the government will rely on a mix of local traffic regulation and clean technology — as opposed to new taxes and bans on the use of cars — to achieve its targets.

France told to put brake on biofuels

Paris. France's planned legislation to curb traffic pollution has been criticized by the French Academy of Sciences, which says that adding oxygenated compounds, such as biofuels, to conventional fuels may create as many problems as it solves.

The addition of these compounds — such as those derived from rapeseed, wheat and beetroot — to fuel would become compulsory in 2000 under legislation to come before the national assembly in the autumn. The proposal, which is a major plank of the legislation, has been enthusiastically supported by the public, ecologists and, not surprisingly, farmers.

But the Green appeal loses some of its shine upon closer examination, says a preliminary academy report on traffic pollution. It argues that, although the addition of such compounds to fuels in the United States has permitted reductions in carbon dioxide and unburnt hydrocarbons, this experience cannot be directly transposed to Europe. This is because of differences in engine designs, cli-

mate and the way cities are built.

Addition of oxygenated derivatives would help to reduce acid pollution, says the report, because the derivatives contain no sulphur, and would cut exhaust levels of benzene, a carcinogen. But it would produce aldehydes that may also be carcinogenic, it says, as well as the nitrogen oxides involved in the formation of ozone. The report contests the wisdom of adding these derivatives to fuel, given the "poorly understood" risks of aldehyde emissions.

The academy concludes that the measure with the most immediate impact on air pollution would be to encourage people to trade in their old vehicles for newer ones fitted with catalytic converters.

The report is also enthusiastic about vehicles running on liquefied gas and electricity. It recommends that gas should be encouraged in fleets of light and heavy goods vehicles and buses, and that greater support should be given to efforts to develop electric vehicles.

Declan Butler

India backing wrong technologies, report says

New Delhi. India needs to reverse its research spending priorities by putting more emphasis on agriculture and biotechnology, and less on defence, nuclear power and space, according to a government report. The report says the current research priorities are out of touch with the government's economic and social goals.

The report, *Technology Priorities for India*, has been released by the Technology Information Forecasting and Assessment Council (TIFAC), a unit of the Department of Science and Technology. It warns that "the balance is tilted so much in favour of military and strategic strength at the expense of economic development that it could imperil national integrity".

Defence, nuclear power and space have eaten up more than half of India's research spending for decades, while health, agricultural and industrial research have received less attention. TIFAC says India's pattern of technology spending is "remarkably similar to that of the Soviet Union just before it broke up".

The report acknowledges that some strategic industries are needed to protect national security. But it argues that spending has ignored the "link between economic development and India's survival as a nation state". And the proportion of resources spent on developing "strategic" technologies is disproportionate to their impact on the economy.

The combined output of the defence, nuclear power and space sectors is only US\$2.5 billion, estimates the report, whereas agriculture accounts for \$71 billion of gross domestic product (GDP) and other industries account for \$31 billion. India's

textile, leather and chemicals industries generate much more money than defence, but spending on research in those areas is one-fifteenth of that spent on defence.

The report points out that India spent \$300 million on nuclear power last year but this generated only 3 per cent of the electricity supply. Thermal power technologies received just \$20 million, although almost three-quarters of India's electricity is produced from coal.

TIFAC urges the government to redefine technology spending priorities in terms of their potential contribution to GDP. According to the study, that would put agricultural research at the top of the list, followed by industrial and health-related research and development, and would leave defence sixth, nuclear power seventh and space ninth. Biotechnology would also be a priority for spending.

In support of this approach, the report argues that, even if India should succeed in developing new armaments, these would be "extremely unlikely" to find export markets. But "even incremental improvements in

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Kalpakham nuclear power station: what price nuclear know-how?

agricultural technology" would make India a major exporter of food, with a large impact on the nation's wealth.

The TIFAC report appears to have cut little ice with the government, however. Yoginder Alagh, the newly appointed science minister, says spending on research in "strategic areas" will need to remain high for reasons of national security. "These are areas in which we cannot get help from any other country," he says. **K. S. Jayaraman**

Ebola bar creates monkey shortage

Tokyo. Japanese scientists say that the rising cost of monkeys and a shortage of supply is hindering research on HIV vaccines and vectors for gene therapy. They blame the problem on a decision made in 1990 by Japan's two major inter-national carriers —

years ago a cynomolgus monkey cost ¥150,000 (US\$1,500) but now it is almost three times that much.

Biomedical researchers say that wild monkeys are more likely to carry viruses and parasites than monkeys bred in captivity for research purposes. The researchers want the government to launch a review of the industry, which they hope will lead to improved supply and better safety standards.

The health ministry acknowledges the potential risks of monkeys imported as pets. One pet-shop owner caught dysentery after importing a monkey, which later died of unknown causes. The ministry may set up a committee to assess the danger and to study biomedical research needs, says an official.

Pharmaceutical companies say that the shortage and high cost of monkeys is creating a bottleneck in the development of gene-therapy vectors and AIDS vaccines, according to Mitsuru Miyata, editor of *Nikkei Biotechnology*, an industry newsletter.

Japan's only clinical trial of a gene therapy last year relied on a US retroviral vector, and on data from the US Food and Drug Administration on its safety and efficacy.

Concern that Japan is falling behind in the race to develop and patent vectors led to the creation last year of a government-backed research and development company, DNAVEC. As well as developing vectors, the company plans to produce methods for testing their safety and efficacy. This cannot be done without a steady supply of monkeys.

Richard Nathan

Nuke ship reincarnated

Tokyo. The controversial nuclear-powered ship *Mutsu* was relaunched last week in Tokyo Bay after a ¥20 billion (US\$200 million) refit as one of the world's largest oceanographic and meteorological research vessels. The 8,600-tonne vessel now belongs to the Japan Marine Science and Technology Centre. It has been stripped of its nuclear reactor and given a diesel-electric hybrid propulsion system. The vessel, renamed *Mirai* ('future'), will have a crew of 80, including 28 researchers.

Japan has invested more than ¥100 billion and nearly 30 years of effort in the vessel. It was launched in 1968 but protests about potential radioactive pollution delayed its maiden voyage by six years.

When it finally set sail its reactor sprang a leak and years of controversy followed as the unwanted vessel was towed from port to port. The controversy brought an end to Japanese ambitions to develop nuclear-powered ships.

R. N.

A dying breed: 'research' monkeys are in short supply in Japan.

Japan Airlines and All Nippon Airways — to stop importing certain species of monkeys to guard against the entry of the Ebola virus.

The decision to stop importing rhesus, cynomolgus and African green monkeys from Africa and South-East Asia for biomedical research was made following international concern about the spread of the virus. The airlines consulted the health ministry but it issued no official restrictions.

Korean Airlines subsequently introduced a similar ban, leaving just one foreign airline in the business of shipping monkeys to Japan. The restricted supply has caused the price of research monkeys to rocket. Five

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

E. Cho

A few (million) dollars more can keep Galileo exploring

Washington. The Jupiter mission of the US Galileo spacecraft may be extended beyond its planned shutdown next year to allow more detailed study of the planet's most intriguing moons, Europa and Io.

Spacecraft managers presented the plan at last week's meeting of the National Aeronautics and Space Administration (NASA) advisory committee for solar system exploration. The idea is still a long way from final approval but it was received enthusiastically by the committee, which is made up of scientists from outside the agency.

Steven Squyres of Cornell University said: "I'm hard pressed to think of a more cost-effective way we could spend our money." This opinion was shared by all the panel members. Marcia Neugebauer of the Jet Propulsion Laboratory (JPL) in Pasadena, California, said "it's stupid to turn off spacecraft" while they are returning useful data.

The Galileo Extended Mission, which was conceived in recent months at JPL, calls for the spacecraft to make between six and eight additional swings past the icy moon Europa in 1998. Galileo is already scheduled to encounter Europa three more times before the end of the Jupiter mission in November 1997.

The extra fly-bys, if approved, would be at closer range, would cover parts of the moon not seen earlier during the mission, and would offer dramatic improvements in image resolution, to between 2 and 20 metres. There would also be a fly-by close to the volcanic moon Io, which was not viewed at close range earlier because of problems with Galileo's tape recorder.

Scientific interest in Europa has been heightened by recent Galileo pictures sug-

gesting that liquid water, or perhaps warm ice, lies beneath the moon's outer crust, as has long been suspected. Those pictures have a resolution of 1.6 kilometres.

Torrence Johnson, Galileo project scientist, said a two-year extended mission could cost as little as \$15 million. To achieve that, the project would streamline its ground



operations and cut its scientific and engineering teams to one-fifth of their present size.

The project would need no additional funding until 1998, he said. That would give NASA time to find the extra money. One option would be to run the extended mission for less than the full two years.

During the period until 1998 the spacecraft's health will be monitored. There is no known problem that would prevent the spacecraft from operating normally through 1999.

Extending missions beyond their planned lifetimes has become a financial problem for the agency, which has postponed until next month a decision on whether to extend several astronomy missions (see *Nature* 382, 193; 1996).

So strapped for cash is the agency's science office that it decided last week to postpone indefinitely a request for proposals for science instruments for the planned Pluto Express mission, which had been loosely targeted for early next century. The reason given was "budgetary uncertainties".

NASA officials say the Pluto mission is not affordable, given a "flat" budget over the next few years, which many believe is the best the agency can hope for.

At last week's advisory committee meeting, Wes Huntress of NASA painted an ironic picture. Despite an "extraordinary year" of scientific and public-relations successes, he said, NASA's science programme was now experiencing the "fastest declining budget of any in the agency".

Tony Reichhardt

NASA aircraft move won't save money, audit finds

San Francisco. A controversial plan to move aircraft used by researchers at NASA sites to one centre will cost \$11.3 million to implement and save very little money, says a NASA audit.

NASA's plan would move research aircraft from sites in Virginia, Ohio, Mississippi and northern California to Dryden Flight Research Center at Edwards Air Force Base in southern California. Daniel Goldin, NASA's administrator, had proposed the move to help cut the agency's budget by \$3 million over the next three years.

Under the plan, all aircraft used for scientific platforms, transport or educational purposes would go to Dryden, where NASA would build a new \$2.8 million computer network system. The plan also involved building helicopter telemetry and radar facilities costing \$990,000 at Ames Research Center in northern California.

NASA's Office of the Inspector General analysed the plan and concluded that these and other costs would rob it of any savings and efficiencies.

Officials in the San Francisco area have fought the move, saying that Ames — which would lose nine research aeroplanes — would suffer most. Representative Anna Eshoo (Democrat, California) is concerned that the loss of most of Ames' aircraft operations would threaten the future of the adjoining Moffett Federal Airfield. She criticized NASA for lack of concern about the impact the move would have on research programmes at Ames.

Ames had been a favoured site for a proposed \$2.5 billion wind-tunnel complex for the development of subsonic airliners, but the plan fell through last year.

NASA officials had estimated that moving the aircraft to one site would save \$23.3 million. The inspector general's report concluded that savings would amount to just \$218,049. The report said: "If the cost of money is factored in, we project that NASA would never recover its financial investment." NASA officials were unavailable for comment.

NASA has managed Moffett as a shared facility for several government agencies and their contractors since the Navy abandoned it in 1994. The Bureau of Alcohol, Tobacco and Firearms, the United States Geological Survey, Lockheed Martin Missiles and Space, the American Red Cross and the Air National Guard are among its tenants. But Moffett has had some trouble lately in its efforts to diversify. A plan to sponsor air cargo carriers who are members of the Civil Reserve Air Fleet at the site has been temporarily scrapped after complaints from local residents.

Sally Lehrman

Mars research boost

Washington. NASA hopes to spend an extra \$1 million next year on the analysis of meteorites believed to have originated on Mars. The space agency expects to solicit research proposals in November, perhaps in cooperation with the National Science Foundation and the Smithsonian Institution, who jointly administer the existing meteorite research programme.

A study team at NASA's Johnson Space Center in Houston has also jumped on the Mars bandwagon, kicking off an investigation into how the agency might conduct a human visit to the planet in the next century. A recent crash exercise to review robotic spacecraft plans for Mars (see *Nature*, 382, 565; 1996) produced only a list of general options — including a sample return in 2003 — with no price tags attached. T.R.

Coordinated plan to protect South China Sea

Hong Kong. An ambitious project to unite bitter political foes in mapping and modeling the environment of the South China Sea is being planned by the Hong Kong University of Science and Technology (HKUST). Scientists at the university say the plan is urgently required to prevent rapid development from overwhelming the sea's ecosystem.

Fifty marine scientists from five countries in the region backed the project at a meeting last December. Last month Professor Chen Jay-chung of HKUST and Gary Heinke, director of the university's institute of environmental studies, took the plan to officials of the United Nations Development Programme (UNDP) in New York. "They were very interested after we told them we had backing from Chinese and Vietnamese officials," says Chen. "They are waiting for our proposal."

Chen says the project would enable builders of a sewage plant in the Philippines, say, to work out the effects of possible sites and treatment systems on the sea as a whole. A similar model is already used in the similarly sized and enclosed Mediterranean Sea, he says.

The first stage of the project would collect data that already exist in various forms in the countries surrounding the sea, and then decide what other data were needed. Facilities and observation stations, including satellites and remote-sensing equipment, would be set up in a second stage.

The full project to set up the model, called Econet, would take up to 15 years and cost US\$150 million. Sponsorship will have to come from the governments involved and the United Nations.

The South China Sea's biodiversity, seabed and the flow patterns of its currents have not been extensively studied, and the sea is not heavily polluted. But that could change quickly, as it is almost fully enclosed by 10 rapidly developing countries.

Proponents of the plan concede that arranging collaboration between these countries will be a formidable political challenge. China, Taiwan, Malaysia, Brunei, the Philippines and Vietnam, for example, have rival claims to a small group of rocks called the Spratly Islands in the centre of the sea, mainly because of oil there.

UN officials were initially sceptical about the project because of the poor political relations between some of the partners. But senior officials in the environmental protection agencies of China and Vietnam are now supportive and the UNDP is interested. It suggested the two-phase approach, and asked for a detailed proposal by the end of the year. Funds for phase one could be available next summer if the project is accepted.

UNDP has also agreed to technical changes that will enable the proposal to be

submitted directly by HKUST as a "regional project," rather than coming through the UK government, which has jurisdiction over Hong Kong only until next year.

The scientists behind the project are rac-



ing against time. Although there is awareness of the need to tackle pollution, action is not catching up with the breakneck speed of coastal development. Growing volumes of industrial waste and untreated domestic sewage are being pumped into the sea.

Development has already taken its toll on what were once large areas of mangroves. Their removal has left formerly protected

coastlines vulnerable to floods.

The project would look for hitherto unknown species in the mangroves, coral reefs and deep ocean mud, and aim to gather full knowledge of the area so that adverse effects of development could be assessed and avoided.

HKUST is drawing up a plan for a three-year first phase, taking in the countries to the north of the sea — Hong Kong, China, Taiwan and Vietnam — and costing \$20 million. But there are already signs of friction between the would-be partners. A \$3.6 million biodiversity study started this year by the six claimants to the Spratlys is in trouble, as China and Vietnam each claims that it has found more species than the other.

Chen admits that China was instrumental in keeping the Philippines out of phase one. "And Vietnam doesn't want to have anything to do with China, they will only deal with us," he says.

He agrees that the players might be more interested in tracking down minerals rather than biological reserves: "There's always a danger of them trying to use the project to their own advantage, but we will have to be aware of that." Heinke concedes that the pace of development could outstrip the ability of the programme to protect the sea. "We have to do it sooner rather than later," he says.

Elisabeth Tacey

Greens attack transgenic plant trials

Munich. Biotechnology companies have conducted field trials of genetically engineered plants in Guatemala without informing local authorities and without taking appropriate biosafety measures, claims a report by Greenpeace, the environmental pressure group.

The pressure group calls on the governments of Guatemala and other Central American countries to declare a moratorium on such trials until mechanisms are put in place to regulate imports of genetically engineered organisms.

Greenpeace began an investigation last year into the activities of several companies with interests in genetically engineered crops. It alleges that the company Asgrow — now a division of Seminis, based at Saticoy, California — conducted trials on transgenic tomatoes, squash and melons without notifying government authorities.

The company also sent genetically engineered seed to Guatemala from the United States by commercial courier service without marking the contents, Greenpeace says.

Greenpeace criticizes a trial on transgenic tomatoes which it says was conducted without systematic application of measures to prevent the accidental release of transgenic material into the environment. It

alleges that greenhouse doors were often left open, for example. This raises the risk of cross-pollination, it argues, which would disturb the ecological balance by increasing the invasiveness and weed potential of the wild population. Tomatoes, like many other economically important crops, are indigenous to the area.

Greenpeace says the lack of legal regulation of these activities is a major problem. Although Asgrow did not transgress any particular law, the pressure group says that the company acted with "a lack of ethics", particularly by not informing authorities of the import and export of gene plasm. If there is damage to the environment, the Guatemala government will not be able to hold companies liable.

Greenpeace says that, as well as stopping field trials immediately, Guatemala should encourage the establishment of a framework agreement in Central America to define policies and standards for the use of genetically engineered products.

It also wants to ban the commercialization of products that present a risk to local flora and fauna. Individual countries should also adopt legislation to control the use of genetically engineered organisms in the region, Greenpeace says. **Alison Abbott**

Last-ditch bid to rescue nuclear test ban treaty

Paris. Supporters of the Comprehensive Test Ban Treaty (CTBT) are preparing a last-ditch attempt to resuscitate it, after almost three years of talks at the Conference on Disarmament in Geneva ended last week in failure to reach consensus.

India complained that the text lacked sufficient commitment to disarmament, and vetoed plans to send the draft treaty to the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASON
REASONS

Gamma

Disarmament would probably have sailed through the assembly, an individually sponsored resolution may need to surmount a steep political hill to get passed.

One fear is that the assembly might decide to amend the current text, upsetting consensus on other contentious issues.

The major stumbling-block to achieving an enforceable treaty seems likely to remain. That is the provision that the treaty cannot become legally binding unless ratified by the five nuclear weapons states — the United States, China, Russia, France and the United Kingdom — and the three 'threshold' weapons states, India, Pakistan and Israel. It was this provision that allowed India to hold the treaty hostage in Geneva.

The provision was included at the demand of China, Russia and the United Kingdom. The United States is said to be flexible on the issue. But, having won over China to supporting a test ban, the United States is now loath to risk a change to the text that might provide China (and perhaps Russia) with a reason to withdraw.

Others hope that India may yet agree to the treaty. One option would be for India to sign and ratify on condition that the weapons states agree to a timetable of nuclear disarmament within five or ten years. If this condition was not met it could withdraw. Instead of holding the treaty hostage, India would be holding hostage the future conduct of the weapons states.

Declan Butler

Nuclear testing: no end in sight.

United Nations General Assembly next month for approval.

Individual countries are now left with the last resort of each sponsoring the current text at the assembly. Australia has offered to take the lead. If the resolution is signed by a large number of countries it could amount to a *de facto* multilateral moratorium on testing, according to Christopher Paine of the Washington-based Natural Resources Defense Council.

There is no reason why a resolution could not go on to become a formal treaty. But, whereas a text backed by the Conference on

South Carolina told to take spent fuel

Washington. A US judge has refused South Carolina's request to block shipments to the state of spent nuclear fuel from foreign research reactors. This clears the way for two shipments of fuel rods from Europe and South America to be sent to storage pools at the Department of Energy (DoE)'s Savannah River site next month.

The state argues that the DoE failed to disclose potential environmental and safety hazards of storing the fuel in the site's underwater basins. It says the department concluded that the basins are not designed for long-term storage of spent fuel.

US District Judge Joseph Anderson rejected the state's request for an emergency injunction to block the shipment of 275 highly-enriched uranium spent fuel rods. A full trial of the state's case is to begin on 25 November.

The Clinton administration wants the fuel, all of which originated in the United States, returned there to prevent the proliferation of nuclear weapons. Secretary of State Warren Christopher said that if South

Carolina's lawsuit were successful it would create "a high risk of release into international commerce of quantities of weapons-grade uranium that could otherwise be brought under US control".

Officials fear that if the United States refuses to accept waste from other countries they will ship it to reprocessing plants in Britain or France to recover highly-enriched uranium, which is usable in weapons.

South Carolina attacked DoE's environmental study of the shipments carried out earlier this year, saying that the department had failed to develop a long-term disposal plan. The state lost a court battle to block initial shipments of spent fuel last year.

The DoE plans to accept 20 tonnes of spent fuel from research reactors in 41 countries over the next 13 years. The 22,700 spent fuel elements in the programme are only a small fraction of DoE's own volume of spent fuel, and are minuscule compared to the quantity of commercial spent fuel in storage at reactor sites across the United States.

David Kramer

NSF backs down on import of Japanese supercomputer

Washington. The US National Science Foundation (NSF) has bowed to pressure from the Congress and the commerce department and suspended approval for the purchase of a US\$35 million Japanese-built supercomputer for use by climate scientists.

Purchase of the NEC computer for use at the National Center for Atmospheric Research (NCAR) at Boulder, Colorado, will be held up for several months while the commerce department and the International Trade Commission establish whether the computer is being "dumped".

Rick Anthes, president of the University Corporation for Atmospheric Research (UCAR), which runs NCAR, says the centre is "developing a number of alternatives on the assumption that the delay could be as much as six months".

NCAR had intended to start using the first of the computer's four processors this autumn, Anthes says. The alternatives being considered include extending existing supercomputer leases, buying time on other organizations' machines and even obtaining new equipment on short-term lease. The NEC machine would have expanded NCAR's computing power by a factor of ten.

Anthes says he understands NSF's decision to suspend approval of the purchase, which was made on 20 August, the day after the commerce department announced its intention to investigate the deal.

Neal Lane, the director of NSF, said in a statement: "I am pleased that the issue of dumping is being properly addressed by the appropriate federal agencies. I am acutely aware that NCAR needs state-of-the-art computational equipment. I feel, however, that acting now on this procurement would be inconsistent with the responsible stewardship of federal funds."

Lane declined to expand on his statement. But the decision to delay the procurement appears to represent a climbdown by NSF and by Jack Gibbons, President Bill Clinton's science adviser, who has said that he supports an open market in supercomputers. NSF officials had said earlier that they were satisfied with UCAR's procedures.

US suppliers have sold several supercomputers to Japanese government agencies in the past two years, but NEC's sale to NCAR would have been its first to the US public sector.

The International Trade Commission and the department will now embark on a lengthy round of decisions and responses to determine whether the sale can go ahead. The first significant milestone will be a "preliminary determination" by the department, promised by January.

Colin Macilwain

Clinton takes historic step to regulate tobacco "as a drug"

Washington. The US Food and Drug Administration (FDA) will begin regulating the tobacco industry under an executive order signed by President Clinton on August 23.

The FDA will implement new rules aimed at curbing smoking by children, based on a finding that nicotine in cigarettes is a drug, and that cigarettes are drug delivery systems. The rules include a ban on vending machine sales, except in adult-only locations, and restrictions on cigarette advertising in youth magazines, on clothing, at sporting events and near schools and playgrounds. Tobacco makers would have to finance a campaign against underage smoking.

In his new book, "Between Hope and History," Clinton writes that the tobacco industry "has no right to peddle cigarettes to children," and called marketing directed at them "immoral." Cigarette workers' unions and the tobacco industry adamantly oppose the changes. □

Einstein medal for Indian chemist

New Delhi. Chintamani Nagesa Ramachandra Rao, a solid-state chemist and president of the Jawaharlal Nehru Centre for Advanced Scientific Research in Bangalore, has been awarded the Einstein Gold Medal by UNESCO. The award recognizes his "outstanding contribution to science".

The medal was first minted in 1979 to mark the 100th anniversary of Einstein's birth. UNESCO's office in New Delhi said that Rao, who has written 700 research papers and 30 books, will receive a diploma and the Einstein medal at a ceremony in Paris later this year. □

NYU sued in 'whistle-blowing' case

Washington. Jan Moor-Jankowski, the former director of New York University's Laboratory for Experimental Medicine and Surgery in Primates (LEMSIP), has filed a lawsuit against the university alleging wrongful dismissal. Moor-Jankowski claims the university fired him in retaliation for his 'whistle-blowing' on crack cocaine experiments conducted on monkeys by another researcher, Ron Wood, last summer (see *Nature* 376, 543; 1996).



The US Department of Agriculture, which ruled in May that Moor-Jankowski had not been victimized by the university, is also named in his suit to the US District Court in New York. □

German venture capital for biotech

Munich. Germany's first venture-capital company dedicated to biotechnology will be launched next month with an initial investment fund of DM75 million (US\$50 million). The company, Global Life Sciences Fund, has three equal partners: Corange, the holding company that owns the pharmaceuticals companies Boehringer-Mannheim and DePuy; Bayerische Vereinbank; and Dutch investment bank, ING. It hopes to raise its fund to DM125 million by attracting further investors. □

Algeria to have atomic commission

Paris. Algeria is to set up an atomic energy commission by the end of the year, its government announced. It will coordinate existing nuclear research, and promote medical, agricultural and industrial applications, said an official from the country's delegation to the International Atomic Energy Agency (IAEA). The

To **save an hour**
each time you
purify histidine-
tagged proteins,
put it on the
tip of a syringe



official added that Algeria has no ambitions to generate nuclear power or build nuclear weapons. Concerns about the latter have eased since Algeria decided last year to adhere to the nuclear non-proliferation treaty. □

\$47 million for nuclear clean-up

Washington. A programme of basic research to support the clean-up of the sites where the United States made its nuclear weapons is to receive its first round of grants from the US Department of Energy (see *Nature* 379, 664; 1996). The department is to award 138 grants worth \$47 million to 52 universities and 11 DOE laboratories. The funds are intended to nurture scientific interest in areas such as glass chemistry and bioremediation, which will eventually be applied to the clean-up programme.

The Pacific Northwest National Laboratory in Washington state did best in the competition, winning ten awards. □

Feds to investigate crop sabotage

Munich. Germany's research minister, Jürgen Rüttgers, has asked the Ministry of the Interior to call in the federal police agency, the Bundeskriminalamt, to locate and prosecute militant activists who destroyed a transgenic crop field experiment earlier this month.

The trial was being carried out by the Max Planck Institute for Breeding Research in Cologne. The aim was to assess possible environmental risks associated with genetically manipulated potatoes. It is the 13th field trial to be destroyed in Germany since January. □

Slap on wrist for NIH

Washington. The National Institutes of Health (NIH) has been fined \$2,500 by the Nuclear Regulatory Commission (NRC) for failing to adequately secure radioactive materials at its research

laboratories in Bethesda, Maryland. The materials were found in unlocked refrigerators last summer, when NRC inspectors went to NIH to investigate an incident in which researchers had ingested radioactive phosphorus-32.

The NRC said it was "particularly concerned" at the number of violations and the fact that similar ones had been discovered earlier in 1995 and the year before. NIH's budget this year is \$12 billion. □

Scientists condemn church raid

Paris. The main trade union representing French researchers, Syndicat National des Chercheurs Scientifiques, protested at the "brutal eviction" from a Paris church of 300 illegal African immigrants. The illegal immigrants, ten of whom were in the 50th day of a hunger strike, were removed by teargas canister-throwing police.

In a letter to Jacques Chirac, the French president, the union said the scientific community was "honoured" to include foreign researchers in its ranks. It said it deplored the way in which new immigration laws were being applied. In many cases, people who have lived legally in France for years have become aliens overnight. □

Mandela names new minister

Cape Town. President Nelson Mandela has appointed Lionel Mtshali, 60, education spokesperson of the Inkatha Freedom Party (IFP), as Minister of Arts, Culture, Science and Technology. He succeeds fellow-IFP MP Ben Ngubane, who party leader Mangosuthu Buthelezi has transferred elsewhere. Mtshali is a former history teacher who was education minister in the Kwazulu homeland from 1990 to 1994. □

We apologize to Anthony Hewish and to Jocelyn Bell Burnell for inadvertently rewriting her role in the pulsar story and for misspelling her name (*Nature* 382, 664; 1996).
Editor, *Nature*.

A column from
the HisTrap™ Kit,
magnified 280%.

Would you like to save one hour in histidine-tagged protein purification?
To load, purify and elute in 15 minutes? Now it's possible—time after time.

HisTrap kit: saving an hour couldn't be simpler

The HisTrap™ Kit presents you with a solution so simple it works with a syringe. Inside the kit is all you'll need for up to ten affinity purification runs: pre-packed cartridges, ready-made stock solutions, a protocol that's easy to follow, and, of course, a syringe.

You'll get high purity recombinant proteins as the kit lets you use the well-known binding of poly(histidine) sequences to immobilized nickel. What's more, it saves you the effort of weighing out buffer salts.

Molecular biologists have been purifying recombinant proteins with Pharmacia Biotech media from the very beginning. Call us at 1 (800) 526 3593 from the USA, +81 (0)3 3492 6949 from Japan or +46 (0)18 16 50 11 from Europe and the rest of the world (or meet us on the Internet at <http://www.biotech.pharmacia.se>). Ask about the kit that saves an hour.



Pharmacia
Biotech

Uppsala, Sweden. (And the rest of the world)



"Purity, simplicity and time savings: a product that gives you all these, lets you take a break from the lab for a bagel, coffee and...hmm, maybe a haircut?" Kyungsun Suh, Ph.D., at the Rockefeller University, New York, NY

Foxy logic on gene pursuit

SIR — Richard M. Lebovitz asks “who else should own the gene but the scientist who first captures it by cloning?” (*Nature* 382, 17; 1996). An ethical answer is: at least in some part the person from whose tissue the gene was cloned. Lebovitz uses an analogy about ownership of a wild fox. Unlike foxes, human genes are not free-ranging but parts of genomes. Summed over the species, these comprise its gene pool. The scientific utility of this concept should not engender confusion between abstraction and physical reality: at any moment, the genes are parts of the bodies of particular people who in free societies retain rights that are governed by the rules of informed consent.

For example, I have donated blood samples in drives to secure matches for bone-marrow transplants and have signed appropriate consent forms. Had the antigenic correspondence been close, I would willingly have donated the tissue required to save a life. However, if I later found that blood from an unsuccessful screening effort had been used for cloning, sequencing and patenting some of my genes without explicit permission, I would feel cheated and might not respond to further appeals, whatever their stated purposes. It is not surprising to find that members of human populations whose genes have been sampled allegedly for scientific purposes would respond with legal challenges once they discovered that their gene sequences might be patented for financial gain by others (*Nature* 381, 11–14; 1996).

Geneticists, presumably advised by lawyers more familiar with the ethics of medicine than blood sports, must draft agreements that specify clearly what rights are being conveyed whenever a tissue sample is taken. Samples and sequences unaccompanied by documentation of genuinely informed consent belong in the same category as other goods of dubious ownership: presumed to have been obtained extralegally and barred from commerce. Antiquities markets provide more relevant precedence here than *Pierson v. Post*.

The hunting parallel fails on other grounds. A fox may be part of nature, but in the United States landowners have a right to say whether anyone may enter onto their lands in its pursuit. Failure to secure permission for hunting on posted lands constitutes trespass. Similarly, oil and gas cannot be sought on private lands without a valid lease; such agreements regularly provide remuneration with initial lease payments and contingent royalties being part of production costs. Why should biotechnology corporations operate differently? In an age of information, scanning genomes for valuable sequences might be more lucrative than drilling lands for oil. The argument that any

given gene sequence probably exists in numerous copies could aid a corporate entity in bargaining with several possessors, but it is likely that supply and demand curves will intersect at some monetary value above zero. Each issue of *Nature* exists in thousands of copies, yet all have value, and the publisher lawfully limits photocopying of the journal's contents.

Oscar Wilde characterized fox-hunting as “the unspeakable in pursuit of the uneatable”. Genetic research belongs on a less reproachable plane.

Robert B. Eckhardt

Pennsylvania State University,
University Park, Pennsylvania 16802, USA
e-mail: eyl@psuvm.psu.edu

Low impact

SIR — In the past decade, the rather peculiar criteria for academic promotion in Italy have attracted international attention. Instances of controversial appointments to tenured positions at the full or associate professor level have been denounced in journals such as *Nature* and the *Lancet*. Losers have quite frequently seemed to have far better scientific credentials than winners, particularly on the basis of citation analysis.

For those not familiar with Italian academic life, it is difficult to know whether such episodes represent the unavoidable drawbacks of an obsolete system of academic promotion (whereby all vacancies are submitted to the Ministry of University in Rome and all appointments are carried out by national committees every 4 to 5 years) or are indicative of more widespread corruption.

An analysis of the performance of Italian science, based on the Institute for Scientific Information (ISI)'s National Science Indicators on Diskette (NSIOD), suggests the latter interpretation. The NSIOD is a database of summary publication and citation statistics reflecting research performance in the sciences by 90 countries between 1981 and 1994. The database contains counts of publications and citations taken from the peer-reviewed journals indexed by ISI.

These data cover 22 fields, and in every case the Italian citation impact is lower than the world citation impact, even in fields of traditional strength and excellence such as physics and chemistry. The mean percentage of fall in impact compared to the world impact is 32 (s.d.=16). If the impact for 1990–94 is compared to that for 1981–85, Italy is losing ground compared to the rest of the world in a number of areas, including 3 of the 5 top fields of Italian science, materials science (–0.15), physics (–0.26) and mathematics (–0.10). More direct compar-

isons can be made within the European Union. Despite a sizeable number of papers published during the period 1981–94 (210,766), the Italian citation impact for all fields (6.34) is higher only than those in Ireland, Portugal, Spain and Greece.

There are several explanations for the current failure of Italian science, including the lack of investment in research and technology by the (many) Italian governments throughout the years. But decades of arbitrary decisions on academic promotion, research appointments (even inclusion in PhD programmes) and funding assignments are likely to exact a heavy toll on the state of science in Italy. The failure of teaching — not more than a third of the students who enrol eventually graduate — is another manifestation of the struggle for power in Italian universities, which disregards scientific merit and genuine academic interest.

Giovanni A. Fava

Department of Psychology,
University of Bologna,
Viale Berti Pichat 5,
I-40127 Bologna, Italy

Innovations catalogue

SIR — Terence Kealey's latest book has received much comment, including that in *Nature* (383, 123–124; 1996), but it begs some questions. If technology is “largely derived from the industrial development of pre-existing technology”, one wonders where did it all begin? With the wheel? Surely Super CRAYS and PCR analysis (for instance) owe their origins at least as much to inquiring minds unfettered by commercial constraints as to the incremental refinement of pre-existing technology back to the year dot?

Second, if Kealey's hypothesis is right, then why is UK manufacturing increasingly keen to collaborate with universities (now up to 31% of companies — *Research Fortnight* 12 June, page 12) and turning away from its own in-house research? Furthermore, the Confederation of British Industry (CBI) is reported as saying that “[s]uch co-operation helps to spread risk and enhance the intellectual and skills pools available to companies”.

Following on from this, the other question Kealey's readers might ask is who is he speaking for? He is speaking for himself certainly, but not for the many scientists who would like to see the UK science base enhanced (as demonstrated by the evidence submitted before the 1993 White Paper on science and technology), nor it would seem, given the CBI view, for industry.

Jonathan Cowie

5 Charlieville Road,
North Heath, Kent DA8 1HJ, UK

Is science losing its objectivity?

John Ziman

The philosophy of science is not independent of the way research is organized. Can scientists produce objective knowledge in a world where their research is increasingly directed towards making money or meeting social needs?

SCIENTISTS know philosophy and sociology as fish know water. They understand instinctively how to live in it without being aware that they are doing so. That is, until the fish bowl is stirred or (horror!) overturned. We seem to be living in just such a time. Science is being shaken up and forced to abandon many of its cherished customs. We need to think hard about what is happening and what we should do, not merely to survive but to serve and delight humanity.

The initial impulse is to defend 'science' against its proclaimed enemies. But much of the pressure comes from equally demanding friends. And what is being defended? Science has already changed a great deal in just a few years. What is the essence that must at all costs be preserved? The bowl is not a black box, but it is filled with an invisible compound of philosophy and sociology. To understand its life-giving properties, this medium needs to be broken down into its component parts, perhaps to be resynthesized in new and more up-to-date forms.

Defining science

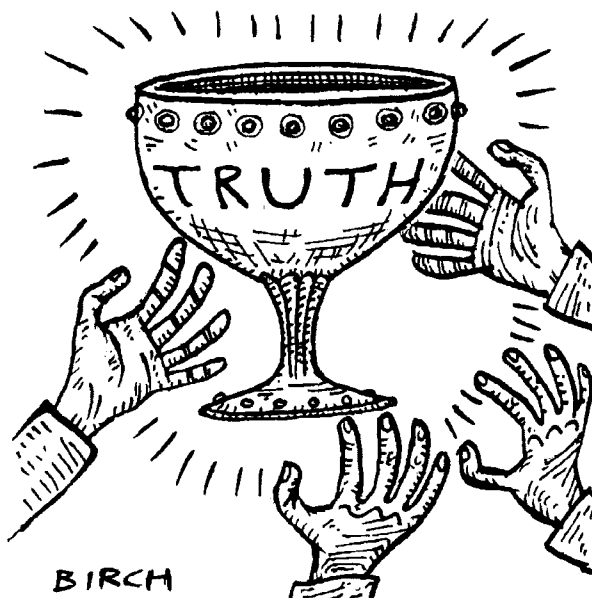
The fundamental question is simple: what is the framework that holds scientists together and keeps their personal rivalries within bounds? The conventional answer is that scientists are united in 'the pursuit of truth'. But some philosophers say that 'truth' is an illusion, whereas others say that it takes many forms, of which only a few are pursued scientifically. Even philosophers of science disagree on just what distinguishes science from other forms of organized knowledge.

What is clear, though, is that the type of knowledge produced by science seems to satisfy certain general principles, such as reliance on observation, explanatory power, universality and objectivity. These principles are abstract and impersonal. They do not tell us what this knowledge is good for, what motivates scientists to seek it or how they should work together in the process. Yet scientists doing basic research have a strong sense of belonging to a community and of being guided in their scientific work by just such principles. How is this achieved?

The answer lies in the fact that what we

mean by basic or 'pure' science can be defined only sociologically. The social institution that has customarily fostered undirected research, without regard for its practical use, is academia. In effect, what we call basic research is almost synonymous with the type of research traditionally carried out in universities¹.

Every PhD student soon learns by



experience that academic science is a distinct human culture, with its own peculiar practices, rules, traditions and conventions. In 1942, Robert Merton² suggested that these practices were governed by a set of unwritten social norms. This type of sociological analysis is now considered to be very questionable, but Merton's norms sum up many familiar social characteristics of academic science in a way that helps to relate them to the philosophical characteristics of scientific knowledge.

The norm of 'communalism', for example, requires that the fruits of research should be regarded as 'public knowledge'. It covers all the practices involved in the communication of research results to other scientists, to students and to society at large. But this has philosophical implications. By insisting on the pooling of personal knowledge gained from individual experience, it stresses the role of observation and experiment in science and underpins scientific realism and empiricism.

'Universality' requires that contributions to science should not be excluded because

of nationality, religion, social status or other irrelevant criteria. In practice, this multicultural meritocratic ideal is achieved very imperfectly. But it does imply that scientific propositions should be general enough to apply in any cultural setting. It encourages scientists to construct abstract theories that claim to explain and unify a wide variety of phenomena.

The idea that academic scientists have to be 'disinterested' means that in presenting their work publicly they must discount any material interests that might prejudice their findings and adopt a humble, neutral, impersonal stance that hides their natural enthusiasm for their own ideas.

'Originality' energizes the scientific enterprise. Academic scientists are expected to be 'self-winding' in their choice of research problems and techniques. Their most cherished traditions celebrate and sustain this important aspect of academic freedom. This norm keeps science progressive and always open to intellectual novelty.

'Scepticism', on the other hand, is the basis for many academic practices, such as critical controversy and peer review. It is not a licence for systematic philosophical doubt, nor for total sociological relativism. But it does stress the systematic testing of research claims in terms of rational qualities such as logical consistency and practical reliability.

The scientific ethos

The close link between social norms and philosophical principles is no accident. It is not even clear which set comes first. It could be argued that the philosophical principles are primary and that the norms sum up the social practices that have naturally developed as scientists have tried to apply these principles in their research. But a sociologist might say that the institutional setting of academic science generates certain practices and that these practices determine the principles regulating the type of knowledge that is produced. The norms and principles are clearly complementary aspects of an ethos whose social and psychological parts are inseparable.

It does not follow, however, that all truth is 'relative' or that scientific know-

edge is 'constructed' entirely to suit certain social 'interests'. All it means is that the progressive unveiling of nature is not a very systematic process. How far we have got in that process — that is, what counts as scientific knowledge at any given moment — is obviously influenced by the way in which research is organized.

This comes out clearly when we consider how academic science is organized. Whatever the formal management structure, academic science is divided into disciplines. That disciplines are usually loosely organized does not make them ineffective. An academic discipline is a global 'invisible college' whose members share a particular research tradition. This is where scientists acquire the theoretical frameworks, codes of practice and technical methods considered to be 'good science'.

Specialization does not stop there. The subdivision of disciplines into narrow research specialities seems to be an unavoidable feature of academic science³. In practice, most academic scientists can satisfy the norms of originality and scepticism only by concentrating for years on what is known, what is hypothesized and what might be feasible in a limited 'problem area'. As a result, basic scientific knowledge is typically fragmented into little islands of near conformity surrounded by interdisciplinary oceans of ignorance. In other words, the philosophical ideal of a unified science is thwarted by institutional and psychological realities.

Motivating factors

The academic ethos says nothing directly about individual motivation or about how academic scientists make a living. Merton himself pointed out that the initial letters of the norms spell out the acronym CUDOS — that is, acclaim or prestige. The assumption is that academic scientists do research and make public their findings in exchange for 'recognition' by their colleagues. Recognition consists of citations in the literature, prizes, medals, titles — and especially employment.

The peculiar feature of academic science is that it developed as an activity engaged in principally by 'academics', whose official employment is to teach rather than to do research. Everybody knows, of course, that university teachers usually owe their posts to their proven research competence and earn further promotion by their research achievements. Still, the convention is that this research is 'their own work', which they are free to carry out and benefit from as individuals.

Paradoxically, academic research developed as the professional occupation of people not specifically paid for doing it. It has always relied on the willingness of universities and other bodies to provide resources for an activity that they do not directly profit from or control.

The key point is that academic science

relies on public and private patronage⁴, in the broadest sense of that old-fashioned word. Society gains directly from the applicable results and from the trained scientists who come out of universities. It also benefits indirectly from the objectivity of basic science. A scientist holding a permanent post as a university teacher is in a position to do 'pure' research, uninfluenced by commercial, political or other external interests.

Forces of change

Academic science is now changing rapidly. Some changes simply reflect scientific and technological progress. As always, the dedication of science to originality is drawing it into novel modes of activity⁵. Individual achievement is being merged into the collective action of multidisciplinary teams. Communication is being speeded up electronically until it

Academic science is undergoing a cultural revolution. It is giving way to 'post-academic' science, which may be so different sociologically and philosophically that it will produce a different type of knowledge

becomes instantly global. Instrumental sophistication is making it easier, but more expensive, to do good science.

More and more forces are pressing on academic science from society at large⁶. In effect, the whole enterprise has now become too large and expensive to be allowed to go its own way. The governments that mainly fund academic research are putting strict financial ceilings on their patronage. The US decision to stop work on the Superconducting Super Collider was the clearest possible signal that this is a worldwide phenomenon. In every country, governments are trying to get better value for their money.

Whatever the cause, there are widespread signs of a decisive break with the academic tradition. This applies to many of the practices associated with Merton's norms, such as conditions of employment, problem choice, criteria of success and other important features. Transition to a 'steady state' regime is imposing on academic science several requirements incompatible with its traditional ethos. Academic science is undergoing a cultural revolution. It is giving way to 'post-academic' science, which may be so different sociologically and philosophically that it will produce a different type of knowledge.

This metamorphosis is still going on. A group of experts on science policy has recently suggested⁷ that the academic mode of knowledge production — what they call 'mode 1' — is being systematically replaced by 'mode 2', a very different

activity. If they are right, what would post-academic science be like?

Privatizing knowledge

The operating philosophy of research will surely stay unchanged. Scientists will go on theorizing and testing their theories by observation and experiment. Indeed, the norm of communalism underlying this attitude will be reinforced by the increasing speed, size and complexity of electronic communication. Novel observations and theories can be discussed in detail with distant colleagues — or even sceptical rivals. National frontiers become irrelevant. Researchers in industrial companies, government laboratories, charitable foundations and universities can work together in the same team. Even the tribal boundaries between disciplines can be disregarded.

But the pressure to open up the interface between academic institutions and industry has an important philosophical effect. Research results that an academic scientist would have published immediately are being identified as 'intellectual property', which may be kept secret for commercial reasons. In other words, post-academic science may no longer be so committed to the principle of 'public knowledge' — traditionally the linchpin of academic science⁷.

Global networks of communication and collaboration would seem to favour social universalism — but not necessarily its philosophical counterpart. Indeed, according to its proponents, the 'new mode of knowledge production' is not directed towards producing knowledge as such: it is directed towards solving specific problems. This undogmatic striving for local understanding is not necessarily deplorable. Post-academic science may no longer be activated by the vision of a unified, universal scientific world picture. But in the end it may prove as effective in closing the gaps in the knowledge map as a single-minded pursuit of general intellectual unity.

Roots in reality

Post-academic science may also line up with many 'postmodern' philosophers in abandoning the age-old attempt to put human understanding on absolutely firm 'foundations'. Some scientists may find it difficult to accept that science should no longer claim that it can provide a universally applicable answer to every problem. I believe this position was always untenable, and that a retreat from it is one of the ways to defuse some of the current public hostility towards 'science'.

This does not mean that post-academic scientists will reject operational realism: they will still construct their accounts of nature on the basis of a firm belief in the existence of an external world whose behaviour is intelligibly regular and not disjoint. On the contrary, although post-academic science will surely still harbour a

considerable amount of research that is not immediately 'useful', this research will be more firmly rooted in the real world. It will draw on, and generate, its problems, techniques and results from all parts of the conventional 'research-and-development spectrum'. Basic research and technological development already inter-penetrate one another: in the long run, they will become inseparable.

Who sets the problems?

Academic science assumes researchers are free, within reasonable limits, to set their own problems. They regard this as the highest form of scientific creativity. By contrast, post-academic scientists will be expected to work together on problems they have not posed personally and to be rewarded principally for their skilful contributions to the success of their team. Competence as a researcher may then count for less than a good record as a team player or as an expert at dealing with certain technical problems.

When resources are limited, even the most basic research does not take place in a power vacuum. It has to be supported financially and administratively by bodies whose interests go beyond the mere production of knowledge. They exercise these interests at the point of maximum leverage — when research problems are being set. All policy talk about foresight, priorities and accountability is really focused on 'problem choice'.

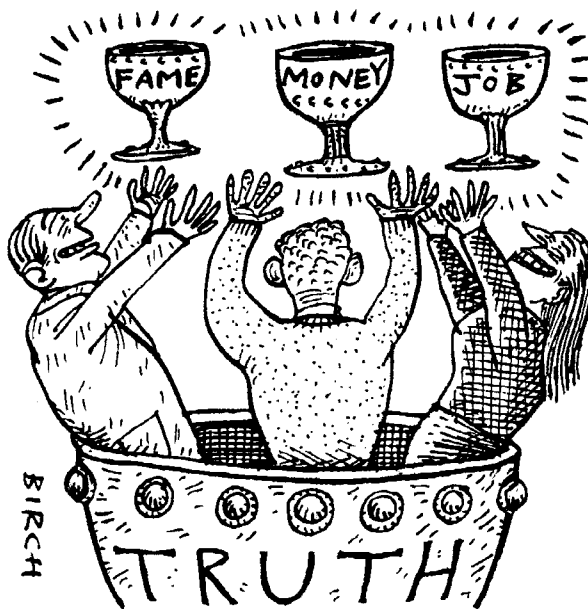
In effect, post-academic science will transform problem choice from an individual to a collective activity³. How well can this be done when the research is not directed towards a practical goal? Academic science has always worked on 'Darwinian' principles³. Scientists do research and offer results on many different problems; knowledge advances in unforeseen directions by the retention of the small proportion of results that survive rigorous testing.

Post-academic science will strive to improve on this apparently wasteful process. It will insist that all research projects are well-designed, non-redundant and directed towards well-posed problems. The unintended effect must be that outlandish projects and wild ideas never get a chance to show their hidden capabilities — which just occasionally open the doors to revolutionary progress.

Celebrating diversity

The world of practice does not carve itself up neatly along the joints between the academic disciplines. In the context of application, all problems require a multi-disciplinary approach. Every important technological development — the transistor, antibiotics, nuclear weapons — com-

bines ideas and techniques from all over the academic map. If we have enough imagination, we can see that this is equally true for research into fundamentals, such as the origin of life or the workings of the brain. The most radical feature of post-academic science could be its unselfconscious pluralism. It will welcome conceptual diversity and not be fearful of possible inconsistencies. If an untidy mixture of theory and practice, computer simulations and numerical data turns out to be the best available solution to a particular problem — so what?



This pragmatism will no longer bar academic science from hybridizing with knowledge and belief systems that do not share the same intellectual values or standards of 'good science'. Specialists from different disciplines, working together as a team, may assemble a mosaic of paradigms, techniques, expertise and practical applications that provides a launch-pad for further advances. But such a mosaic may not be very stable and may not have been built on sound intellectual principles and firm institutional soil. There may not be anybody waiting to welcome the astronauts when they parachute back to Earth.

In harsh reality, moreover, practical problems seldom appear out of nowhere, without antecedents. The world where research is to be applied is already highly structured: problems to be tackled are normally set and funded by their existing organizational 'owners', such as industrial companies, government departments or health services. In the effort to overcome the academic vice of narrow specialization, post-academic science may find that it has put itself in the hands of bodies even more parochial, fragmented and restrictive than the disciplines from which it has escaped. It may become even more difficult to begin research on a problem not

already on the agenda of a wealthy funding body — a serious issue in a world where not all socially important problems are of recognized commercial, technological or political concern.

The price of excellence

Again, post-academic science will distrust the élitism of peer review and replace or bolster it with quality control of people, projects and performance. But this usually entails a much broader notion of 'excellence' than the traditional academic criteria for 'good science'. So greater importance may be attached to entrepreneurial and managerial skills, such as the ability to oversee the larger cycles of action in the research system.

The trouble is that the test of practical utility does not operate in basic research, where organized scepticism² is the only real protection against persistent error. The gravest threat to the reliability of scientific knowledge could be obsessive monitoring of the accountability and performance of researchers at the expense of systematic intellectual criticism of their claims. In any case, considerable intellectual uncertainty is inevitable in areas where post-academic science becomes entangled with 'trans-epistemic' issues, such as questions over bovine spongiform encephalopathy where 'nonscientific' social, environmental and humanistic values are involved.

Who pays the pipers?

Mode 2 researchers, we are told, work in shifting teams, like small companies producing goods for a competitive market. Jobs are never secure. As teams reorganize to tackle new problems, some researchers have to move elsewhere to make room for new people with new skills. As a result, few individuals have stable opportunities to establish or exercise their expertise. The contrast with tenured posts in academic institutions could not be greater.

But it is unrealistic to suppose that today's system of multifunctional universities will give way to a genuine market system sustaining many small, commercially independent research enterprises. Researchers will mostly remain full-time employees of universities, government laboratories, charitable foundations or industrial companies. If not, they simply won't be able to invest in the facilities they need.

Although post-academic science may look attractively unbureaucratic, it will really be heavily capital-intensive. It will continue to be funded and managed by a complex of governmental bodies, large public institutions and private corporations. The same questions will still be

asked: who will pay the pipers and what tunes should they be called on to play?

Industrial concerns

It seems to me, in fact, that many of the suggested differences between mode 1 and 2 are not changes from an old to a new mode of knowledge production. They typify the long-established distinction between 'pure' and 'applied' research, a distinction institutionalized nearly a century ago. There has always been a cultural gap between 'academic' science in universities and 'industrial' science in industrial laboratories⁹. In reality, the former was never totally pure, nor the latter entirely utilitarian, but they were organized differently.

Mode 2 actually reads like a 'post-industrial' version of applied science. Industrial laboratories, like industrial companies, used to be large, monolithic organizations run from the top by a hierarchy of managers. In the post-industrial era (so many economists and business experts tell us) market competition will replace command management. Even in large multinational corporations, global networks, profit centres and independent contractors will replace management charts, directorates and service departments. Naturally enough, technological research and development in the private sector is being reorganized along similar lines. Yet it is still directed towards the same objectives — primarily, financial profit — and is subject to the same socioeconomic imperatives.

What might really be happening is that the evolution of industrial science into post-industrial science is closing the gap between pure and applied research. Several factors are working towards a single post-academic culture. Scientific developments are blurring the distinction between fundamental and exploitable discoveries. Technological developments are generating heterogeneous hybrid teams that override institutional loyalties. Economic conditions are forcing the two cultures into the same organizational mould.

Indeed, a deliberate effort at a high level of political and managerial authority would probably now be required to keep the two systems from coalescing in style and function. But such a merger not only raises many practical issues of funding, disciplinary identity, criteria of excellence, career aspirations, intellectual property rights, institutional management and so on. It also brings face to face two very different sets of structural principles. In this confrontation, mode 2 seems already to be ousting mode 1. The academic ethos may well survive as an attractive but somewhat dated ideology; but the effective culture of post-academic science may well be predominantly post-industrial.

Who's afraid of postmodernism?

The philosophical background to science is often simply taken for granted. I have tried

to show that it is closely connected with the way in which research is organized and carried out. Changes in the social framework of science eventually lead to changes in its philosophical principles. Indeed, a cultural theorist might argue that the transition to post-academic science, organized along post-industrial lines, is necessarily reflected in a transition from a 'modern' to a more 'postmodern' philosophical stance¹⁰.

I do not accept this necessity. I am not suggesting that science is abandoning all its norms and principles and 'going post-modern' in some fashionable pseudo-intellectual sense. It is nonsense, for example, to say that today 'anything goes'. Most of the likely changes are mild, even benign. Some are much-needed corrections to the excesses of 'scientism'. Others are welcome antidotes to the extreme rationalism that has long plagued the philosophy of science. It must be a good thing to help rescue the scientific imagination from entrenched specialization. And localized pragmatism will largely compensate for the fragmentation of theoretical standards of scientific validity. Some of these changes may be philosophical dynamite, but they scarcely affect work at the coalface of research, and I doubt whether scientists or their leaders will even notice them.

Incorporating interests

But there is a serious threat to one fundamental feature of academic science — its objectivity. Objectivity can never be absolute. Philosophers and sociologists agree that the notion of a truly objective disinterested 'seeker after truth' is incompatible with the realities of social existence. We all have personal interests and institutional values that we are bound to promote in our scientific work, however hard we try to suppress them. The virtue of academic science was that it took a strong line in support of the norm of 'disinterestedness' and often managed in practice almost to live up to its ideals.

The transition to post-academic science is eroding the practices that underpin this norm. 'Public knowledge' is being transformed into 'intellectual property'. Basic research networks include many research groups with direct industrial interests. Researchers will not be protected from commercial influences by academic tenure. Their work will often deal with matters where social values — safety, profitability, efficacy — must have the highest priority.

In general, then, post-academic science is bound to be shot through with social interests. It will surely defend objectivity as an ideal, impossible to realize completely in practice but always to be respected and desired. But if all research arises in close connection with its potential for application, there may never be any occasions where this ideal is paramount — where conscience insists we must publicly voice Galileo's whisper: "It really does move!"

Scientific objectivity is not an abstract philosophical virtue. It is a cultural norm embodied in a web of social practices. Academic scientists incorporate the norm of 'disinterestedness' into their system of personal values as a result of their own experience in research situations where these practices are systematically observed. They never forget the moment when they were reproved by their professor for failing to present fairly the arguments of their opponents or for conveniently 'forgetting' an awkward fact that invalidates their exciting discovery. From then on, they know in their hearts that this is how real scientists ought to behave. It is hard to see how this norm will be sustained when there are few situations yielding the relevant experience.

Objectivity is what makes science so valuable in society. It is the public guarantee of reliable disinterested knowledge. Science plays a unique role in settling factual disputes. This is not because it is particularly rational or because it necessarily embodies the truth: it is because it has a well-deserved reputation for impartiality on material issues. The complex fabric of democratic society is held together by trust in this objectivity, exercised openly by scientific experts. Without science as an independent arbiter, many social conflicts could be resolved only by reference to political authority or by a direct appeal to force¹¹.

Science and technology have always been growth enterprises. Scientists look forward to ever-more exciting achievements. But when we reflect on the prospects for change, we should not fall into a reckless, purely celebratory mood. My fears may be exaggerated. Other factors may be at work, completely altering the situation. Is there anything that should be done about it — and if so, what? □

John Ziman, emeritus professor of physics at the University of Bristol, is at 27 Little London Green, Oakley, Aylesbury, Buckinghamshire HP18 9QL, UK.

This article is based on the 1995 Medawar lecture of the Royal Society, given by the author in London on 29 June and in Edinburgh on 27 September 1995.

1. Ziman, J. M. *Basically, It's Purely Academic* (in the press).
2. Merton, R. K. *The Sociology of Science* 267–268 (Chicago University Press, Chicago, 1973).
3. Ziman, J. M. *Of One Mind: The Collectivization of Science* 99–120; 265–310 (American Institute of Physics, New York, 1995).
4. Turner, S. P. In *Theories of Science and Society* (eds Cozzens, S. E. & Gieryn, T. F.) 185–211 (Indiana University Press, Bloomington, 1990).
5. Gibbons, M. et al. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies* (Sage, London, 1994).
6. Ziman, J. M. *Prometheus Bound: Science in a Dynamic Steady State* (Cambridge University Press, Cambridge, 1994).
7. Ziman, J. M. *Public Knowledge: The Social Dimension of Science* (Cambridge University Press, Cambridge, 1968).
8. Campbell, D. T. In *The Philosophy of Karl Popper* (ed. Schilpp, P. A.) 413–463 (Open Court, La Salle, IL, 1974).
9. Ziman, J. M. *An Introduction to Science Studies* (Cambridge University Press, Cambridge, 1984).
10. Toulmin, S. *Cosmopolis: The Hidden Agenda of Modernity* (University of Chicago Press, Chicago, 1992).
11. Ezrahi, Y. *The Descent of Icarus: Science and the Transformation of Contemporary Democracy* (Harvard University Press, Cambridge, MA, 1990).

Sacred cows, science and uncertainties

David C. G. Skegg

An analysis of the transmission dynamics of bovine spongiform encephalopathy (BSE) in Britain provides options for cattle-culling policy. Deciding which option to implement is another matter.

EPIDEMICS are often poorly understood until they have nearly run their course. As recently as 1991, the UK Ministry of Agriculture, Fisheries and Food (MAFF) forecast that BSE would be virtually eliminated from British cattle by 1995 (ref. 1). There were actually about 14,000 new cases in 1995, so the MAFF prediction was one of many optimistic assumptions that have characterized the BSE epidemic. Another was that maternal transmission from cow to calf occurs either rarely or not at all.

The latter assumption was shattered by the announcement earlier this month (1 August) of interim results from a study launched in 1989. This compared the incidence of BSE in two cohorts of animals: offspring of cows that developed BSE, and a carefully matched control group whose mothers did not. A report has been submitted for publication², but findings already released suggest that maternal transmission occurred from about 10% of the infected cows whose calves were studied. This result challenges many of our ideas about BSE, but the first question is what influence maternal transmission will have on the future course of the epidemic. There is compelling evidence that the vast majority of BSE cases have resulted from contaminated feed, but will maternal transmission prevent or delay the eradication of the disease? A simple-minded assumption would be that the predicted number of cases over the next few years must now be increased substantially in the light of maternal transmission. Paradoxically, this may not be so.

This conclusion emerges from an analysis of the transmission dynamics of BSE by Anderson *et al.*, published on page 779 of this issue³. Five months ago, Gore⁴ rightly pointed out that previous projections of the BSE epidemic had not been informed by modern developments in mathematical epidemiology. Anderson and his colleagues have now answered this challenge by developing statistical models to estimate trends in the incidence of new infections (rather than just new cases of clinical BSE). They used the technique of back-calculation, which was developed to predict the course of the AIDS epidemic⁵. Their models take account of key variables

such as the incubation period of BSE, exposure and susceptibility to the infection at different ages, and demographic features of the British cattle population. The new information about maternal transmission is also incorporated. The arcane mathematics should not blind anyone to the limitations of the basic data, because many variables have to be estimated from a single age- and time-specific data set.

IMAGE
UNIMAGINABLE
FOR A COPIABLE
FOR A RIGHT
REASONS

Incineration — the end of the line for BSE-infected cattle.

The whole exercise requires assumptions about certain variables, which affect the results to a greater or lesser extent.

One of the virtues of statistical modelling is that it identifies key gaps in knowledge. In the case of BSE, such gaps are legion. They include the distributions of ages at infection and of incubation periods, and the extent to which cattle at different stages of the incubation period are infectious to other cattle — both through contaminated feed and maternal transmission.

The analysis by Anderson *et al.* suggests that the feed ban implemented in late 1988 had an immediate and lasting impact (see box overleaf for a chronology of the BSE epidemic). Infection through contaminated feed is predicted to have continued, although at a much lower level, until it became negligible by the middle of 1994. Since that time, virtually all new infections would have occurred by the maternal route. The model that best fits the data assumes that the 10% maternal transmission rate applies for the last half-year of the incubation period. Using this preferred model, the authors predict that there will be about 340 new infections and 14,000 new cases of BSE during the six

years from 1996 to 2001. In the absence of maternal transmission, there would be no new infections but a predicted 20,000 new cases over the same period. The lower number of cases with maternal transmission must reflect a lower predicted number of feed-borne cases that are still in the incubation period. It would be interesting to know whether the authors regard this as a real effect, a chance finding or an artefact of the model.

Maternal transmission alone cannot sustain the epidemic of BSE in British cattle, which is predicted to fade close to extinction by the year 2001. Clearly, many more cases will occur in the interim, and there is considerable statistical uncertainty in the timetable for extinction. Thus the preferred model predicts 221 new cases in the year 2000, but the 95% prediction interval ranges from 128 to 3,660 cases.

One of the most valuable features of the modelling approach is that it enables prediction of the impact of different culling policies.

In the past, British veterinary surgeons have objected that proposals to slaughter animals have had no more scientific justification than the burning of witches. Anderson *et al.* estimate, for various culling policies, the number of cases of BSE avoided, the total number of cattle to be sacrificed and the number of herds affected (see table on page 787). Sadly it is not possible to make a big dent in the expected number of cases without slaughtering an enormous number of animals; for example, a combined herd-targeted and maternally targeted policy (No. 11 in the table) is predicted to save about 1,490 (21%) out of 6,950 projected cases of BSE, but could involve the slaughter of up to 44,000 cattle. The UK government will want to review its current policy in the light of the new analysis, especially with respect to maternal transmission. The selection of a culling policy depends on complex political, economic, public health and ethical considerations. People should insist that the objectives of any policy are clearly defined.

Another feature of the transmission dynamics that requires further work is the implication for human risk assessment. The main reason for our intense concern about BSE is the suggestion that it might

Nigel Dickinson/Still Pictures

Chronology of the BSE epidemic

1985 (April)	BSE first observed clinically
1986 (November)	Disease identified as a spongiform encephalopathy
1987 (April)	Initial epidemiological studies started
1987 (December)	Initial epidemiological studies incriminate ruminant-derived meat and bone meal as a cause of BSE
1988 (July)	Ruminant feed ban introduced
1989 (November)	Ban on use of certain specified bovine offals for human consumption in England and Wales
1992	BSE epidemic peaks at 36,681 cases for the year
1993 (July)	Announcement of 100,000th confirmed case of BSE
1996 (March)	Announcement of suspected link between BSE and a new variant of human Creutzfeldt-Jakob disease

be linked to the occurrence of a new variant of human Creutzfeldt-Jakob disease (vCJD)⁶. Anderson *et al.* estimate that about 446,000 infected animals entered the human food chain before the ban on specified bovine offals at the end of 1989 (see box), with another 283,000 before the end of 1995. It is too early to be sure whether or not this endangered human health, and whether or not the small number of cases of vCJD so far detected foreshadow a larger epidemic. But the approach pioneered here will assist in making detailed estimates of human consumption of products derived from BSE-infected cattle at various stages of the incubation period in different calendar years.

Presumably the modest reduction in the projected number of cases of BSE explains why the chairman of the Spongiform Encephalopathy Advisory Committee is reported as describing the new information about maternal transmission as "potentially good news"⁷. Other scientists may not be so sanguine. Not only is it disturbing that transmission at a substantial level has been detected so belatedly, but the discovery also calls into question many sacred cows about BSE. It is widely held that the agent causing BSE is not present in blood or most other tissues. Does maternal transmission occur before, during or after birth? Is it via the embryo, the placenta, milk or what? None of these alternatives is palatable for those who must advise the public. Knowledge of scrapie, an analogous disease of sheep, might suggest the placenta as the most likely culprit⁸. But inoculation of bovine placenta (or blood) into mice has shown no detectable infectivity⁹. This implies that the mouse bioassay may be far too insensitive, and raises questions about

the safety of other bovine tissues apart from the brain and spinal cord.

The continual retreat from entrenched positions about BSE has damaged the credibility of science as well as of politicians. For instance, the finding of a 10% maternal transmission rate may be seen as conflicting with the results of a previous epidemiological case-control study¹⁰, which were widely interpreted as negative. In fact the results are not conflicting: the previous study was too small to have adequate statistical power, and the authors pointed out that their findings were consistent with "rates of maternal transmission of between 0 and 13 per cent"¹⁰.

Reviewing the epidemiological research on BSE¹¹

leaves one with an impression of solid work by too few scientists, involving too few studies and too few animals. The portfolio of laboratory research has been

no more timely or adequate. According to available figures, during the three financial years from 1988 to 1991 the UK government invested the paltry sum of £3 million (about US\$5 million) on BSE-related research¹¹. It is easy to be wise with hindsight, but the BSE saga should be a warning to all countries not to skimp on research into emerging problems that have potentially serious consequences for agriculture or public health. □

David C. G. Skegg is in the Department of Preventive and Social Medicine, University of Otago, PO Box 913, Dunedin, New Zealand.

1. Bown, W. *New Scientist*, **10**, 6 April 1991.
2. Wilesmith, J. W., Wells, G. A. H., Ryan, J. B. M., Gavier-Widen, D. & Simmons, M. M. *Nature* (submitted).
3. Anderson, R. M. *et al. Nature* **382**, 779-788 (1996).
4. Gore, S. M. *Br. Med. J.* **312**, 791-793 (1996).
5. Brookmeyer, R. & Gail, M. H. *Lancet* **ii**, 1320-1322 (1986).
6. Will, R. G. *et al. Lancet* **347**, 921-925 (1996).
7. Bradbury, J. *Lancet* **348**, 393 (1996).
8. Detwiler, L. A. *Rev. Sci. Tech. Off. Int. Epiz.* **11**, 491-537 (1992).
9. *Bovine Spongiform Encephalopathy in Great Britain: A Progress Report* (Ministry of Agriculture, Fisheries & Food, London, May 1996).
10. Hoinville, L. J., Wilesmith, J. W. & Richards, M. S. *Vet. Rec.* **136**, 312-318 (1995).
11. *Programme to Eradicate BSE in the United Kingdom* (Ministry of Agriculture, Fisheries & Food, London, May 1996).

REGULATORY BIOLOGY

Competition for brain space

Jared M. Diamond

EYES have long been regarded as the ultimate test for theories of natural selection: could the gradual accumulation of small random mutations really give rise to such a complex, precisely integrated system¹? But eyes also exemplify an opposite phenomenon: the evolutionary atrophy of eyes in animals living in darkness, and of disused biological structures in general. That phenomenon is in turn linked to an even bigger problem, the evolutionary pruning and quantitative fine-tuning of reserve capacities that pervades regulatory biology. The factors behind eye atrophy in blind mole rats have been illuminated by Eviatar Nevo and his colleagues in a series of elegant studies²⁻¹¹, the significance of which extends far beyond those remarkable animals.

Despite creationists' assertions about the unlikelihood of eyes arising through natural selection, photoreceptors have emerged independently in many animal lineages¹². Evolutionary reduction of eyes in animals no longer needing them proves to be even more frequent, having been documented in hundreds of species belonging to 50 genera and 11 families of subterranean mammals alone, ranging from insectivores (moles) through rodents (mole rats) to marsupials (the marsupial

mole). Blind cave-dwelling and subterranean fish, amphibia, reptiles and invertebrates are legion, while above-ground examples occur among echo-locating microchiropteran bats, aquatic shrews and river dolphins. Outside the visual system, disuse has led to atrophy of olfaction in cetaceans, the appendix in humans, tails in apes, vitamin C synthesis in primates, and substrate synthesis by mutant bacteria (auxotrophic mutants) grown in substrate-containing media.

Two theories to explain such evolutionary reductions of disused structures were formulated by Darwin and are still debated today. According to one of them, re-expressed in modern genetic terms, deleterious mutations of genes coding for any structure arise constantly and are eliminated by natural selection if the structure is used but are no longer eliminated if the structure falls into disuse. That undoubted fact is reflected in the greatly increased rate of accumulation of mutations in silent pseudogenes compared to the corresponding coding genes. The second theory instead hypothesizes a positive selective advantage to getting rid of a disused structure because of thereby eliminating some associated cost — inviting the question of what that cost might

be. Analogous theories for why derelict buildings disappear are that they eventually collapse because of the gradual accumulation of rust and decay left unrepaired, or that the owner actively tears them down to make more profitable use of the land.

Blind mole rats (genus *Spalax*) carry eye reduction to its extreme among mammals. They spend most of their lives in total darkness in underground runways, where eyes would be useless. In fact, their eyes are reduced to less than 1 mm in diameter, 100 times smaller than those of above-ground rodents. The *Spalax* eye contains a mere 823 retinal ganglion cells (in contrast to the 100,000 in a hamster eye) and supplies information to fewer than 1,000 optic nerve fibres. As if that weren't enough, each eye is covered permanently with thick skin and fur, lies embedded in a large gland, and lacks a pupil, functional lens or eye muscles. So it is incapable of forming an image and could at most distinguish between darkness and a diffuse blur of light. Not surprisingly, mole rats are functionally blind: a light flash yields no behavioural response of the animal and no evoked electrophysiological response in its brain.

But *Spalax* does exploit its capacity for detecting diffuse light/dark differences, because it exhibits rhythms of circadian activity entrained by light. Removing the eyes abolishes that photoperiodic response and disturbs the animal's thermoregulation. So atrophy of the visual system proves to be highly selective: although the dorsal lateral geniculate nucleus, retina and other structures associated with image formation are greatly reduced, there is no atrophy of the suprachiasmatic nucleus, the site of the circadian pacemaker involved in photoperiodism and reception of diffuse light. That is, the visual system has undergone a careful dismantling, not blows of the wrecking ball.

What selective factors have driven this pruning of the visual system? The studies of Nevo and colleagues identify savings of two large costs, both of which can serve as models for factors driving evolutionary reductions of other disused structures. One saved cost is metabolic energy. Retinal oxygen consumption per gram is among the highest of any tissue (100 times the average value for the body as a whole). Brain metabolism is also high, accounting for 20% of a small mammal's oxygen and glucose consumption. Because the visual cortex has nearly the highest oxygen consumption of any part of the brain, and accounts for 10% of the brain volume of a typical sighted mammal, it must be the site of at least that percentage of brain metabolism. A quick calculation then shows that, by eliminating most of the brain's visual structures, consuming at least 10% of 20% of whole-animal oxygen intake, *Spalax* saves at least 2% of its en-

ergy budget. In the evolutionary marketplace, that is a potent selective force.

But Nevo and colleagues have identified another, possibly even more important, saved cost. *Spalax* has better things to do with its brain space than waste it on non-functional visual cortex. As a blind animal, it relies on exquisite development of other sensory modalities: vibrissae and bristles on its face, and body touch sense generally, as it feels its way while using its teeth and head to dig a burrow; olfaction, to detect odours in other *Spalax* individuals' urine and possibly in edible parts of plants; sound (male and female communicate noisily once the male has dug its way into a receptive female's burrow); seismic signals (they communicate between burrows by thumping with the head on the burrow roof in a species-specific signature



Out of sight — such eyes as the blind mole rat possesses are covered with skin and fur.

pattern, like Morse code); and magnetic compass orientation, by which each *Spalax* finds its way back to its nest through a labyrinthine tunnel system.

All of those sensory modalities demand brain processing space. *Spalax*'s somatosensory cortex is 70% larger than that of a sighted rat of the same body size, and somatosensory processing takes over the entire volume of occipital cortex that a rat would dedicate to visual cortex. That is, disposing of the useless surplus in the visual system frees a mere 2% of *Spalax*'s energy budget, but 10% of its brain volume, to devote to other purposes.

This spatial reasoning points to the solution to a problem in regulatory biology. Many disused structures that have become reduced or lost through natural selection, such as individual enzymes, are initially so small that the saving of their contribution to an animal's total energy budget must be utterly insignificant as a selective factor. This conclusion applies with even more force to quantitative evolutionary fine-tuning of reserve capacity than to elimination or drastic reduction. Why is the synthesis of so many enzymes and transporters genetically programmed such that inter-individual variation of activity is less than 25% (ref. 13)? It's easy to see why 25% too little of an enzyme is disadvantageous, but why should 25% too much be equally disadvantageous? For example, the activity of the vertebrate

intestinal brush-border sodium glucose transporter (SGLT1) varies intraspecifically between individuals by only 25%, but varies 1,000-fold between species according to the glucose content of their natural diet (high in herbivores, low in carnivores). At most, that glucose transporter accounts for only 1% of brush-border membrane protein and less than 0.01% of total body protein.

Having a 25% excess of such a minor component clearly doesn't constitute an encumbrance on an animal's metabolism. But think of SGLT1 competing for limited brush-border membrane space, just as the visual system competes for limited space in *Spalax*'s brain. Although SGLT1 may constitute less than 0.01% of total body protein, it may be 10% of intestinal brush-border membrane-spanning transporter protein. The total brush-border content of all such transporters cannot be increased without interfering with brush-border membrane function, which depends on the membrane consisting mostly of a lipid bilayer. A small excess of glucose transporter incurs a lost-opportunity cost: prune it by 25%, and the membrane can accommodate that much more of transporters for amino acids, vitamins or other nutrients. It's easy to see how 25% more amino-acid

absorption could bring as big advantages as 10% more brain space for functional sensory modalities.

I am thereby reminded of a principle associated with the name of the comparative physiologist August Krogh: for every biological problem, there exists some animal species best suited to solving it. Through the work of Nevo and colleagues, the eyes of blind mole rats now join a select company of animal models that includes squid axons, pigeon breast-muscle enzymes and developing fruitflies. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024-1751, USA.

1. Dawkins, R. *Nature* **368**, 690–691 (1994).
2. Ben-Shlomo, R., Ritte, U. & Nevo, E. *Behav. Genet.* **26**, 177–184 (1996).
3. Ben-Shlomo, R., Ritte, U. & Nevo, E. *Behav. Genet.* **25**, 239–245 (1995).
4. Nevo, E. *Acta Theoretologica Suppl.* **3**, 9–31 (1995).
5. Nevo, E. *Proc. Zool. Soc. (Calcutta), Haldane Comm. Vol.*, 27–40 (1993).
6. Argamaso, S. M. et al. *Biophys. Chem.* **56**, 3–11 (1995).
7. Shanas, U. et al. *J. Zool. (Lond.)* **237**, 195–210 (1995).
8. Cooper, H. M., Herbin, M. & Nevo, E. *Nature* **361**, 156–159 (1993).
9. Rehkaemper, G., Necker, R. & Nevo, E. *J. Comp. Neurol.* **347**, 570–584 (1994).
10. Nevo, E. in *Diversity in Biological Design* (eds Weibel, E. et al.) (Cambridge Univ. Press, in the press).
11. Nevo, E. *Circadian Light Reception and Regulation, Human Frontier Science Program* (Lyon, 1996).
12. Salvini-Plawen, L. V. & Mayr, E. *Evol. Biol.* **10**, 207–263 (1977).
13. Karasov, W. & Diamond, J. *J. Comp. Physiol.* **152**, 105–116 (1983).

Scum of the Earth after all

P. Jonathan Patchett

WHERE do continents come from? On page 773 of this issue¹, Stein and Goldstein discuss the origin of ancient crust in Africa and Arabia that helps to answer this question. It appears that large submarine plateaux, like the present-day Ontong Java Plateau, may have formed the basis for an appreciable amount of our continental crust.

An important underlying puzzle is why the accumulation of continental crust at the surface of the Earth (part *a* of the figure) does not reflect the Earth's thermal history very well. Part *b* of the figure shows the heat produced by the decay of long-lived radioisotopes. That heat production, the main driving force for mantle movements and plate tectonics, has declined by a factor of five since the birth of the Earth. The continental crust, sweated out of the mantle during these heat-driven processes, is the only long-lived product of that activity. So why does its growth not reflect the thermal history more closely?

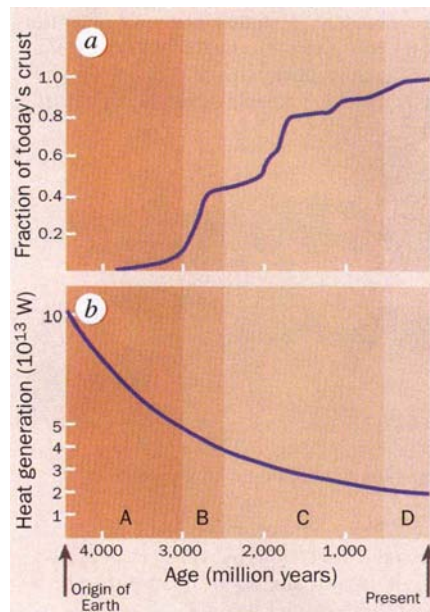
Almost all the Earth's heat loss takes place either by conduction through plates to the oceans or atmosphere, or at the large mid-ocean ridges where oceanic crust is continually being made. The cold ocean crust falls back into the mantle again at subduction zones, and continental crust is only produced because this subducting ocean crust has absorbed water. When this water, or the small proportion of it that can be trapped in higher-pressure minerals, is taken down below about 150 km, it reduces the melting temperature of the mantle by more than 400 °C. As a result, magmas are produced that form volcanic chains or 'arcs', either in the ocean, such as the Tonga Islands, or on the edges of continents, such as the Andes Mountains. These magmas are often less dense and more silica-rich (more like the composition of the continents) than those produced at mid-ocean ridges. Traditionally, continent growth has been explained as the accumulation of these arc magmas, either directly or when island arcs are scraped off subducting ocean plates onto the edges of continents.

Of the ocean crust and mantle rock masses involved, only a tiny proportion melts to produce arcs — the production of arcs and continents is only a by-product of the continual recycling of oceanic plates. The heat budget does not require continents to be made at all! So the continents are quite incidental to the main processes of the solid Earth, like scum on the surface of a liquid^{2,3}.

We might still expect the continent growth curve to mirror the heat generation record, in the sense that more heat implies more ocean crust cycling, more

arcs at subduction sites, and therefore more continental crust. One might also argue that continent growth depends so heavily on the ocean, which sedimentary rocks tell us was present long before three billion years ago, that a continuing slow dribble of continental growth would be expected. But the actual growth of a typical continent isn't at all smooth.

Before three billion years ago (segment



A heat engine for continental growth? *a*, Cumulative crustal growth¹² for the best-studied region, North America and Europe, which were a single land mass for most of the Proterozoic (segment C). Note the major pulse of growth in the late Archaean (B) and later pulses in the Proterozoic. Africa/Arabia differs in the timing, but not the magnitude, of Proterozoic growth episodes, including the pulse 900–600 Myr ago discussed by Stein and Goldstein¹. *b*, Evolution of radioactive heating in the Earth, principally due to ⁴⁰K, ²³⁵U and ²³⁸U, from 4,600 Myr ago to the present.

A of the figure), very little continental crust had accumulated. We have to suppose that the early Earth was either too tectonically active to allow arcs to form and stick together, or that the system was just as efficient at destroying continental crust as making it. Then a sudden spurt of growth occurred (segment B) that is represented in all present-day continents. Later, during the Proterozoic (segment C), the typical continent underwent two or three more major addition episodes, and in the Phanerozoic (segment D), some continents have grown, such as the North America/Europe of the figure, but others, such as Africa/Arabia, have not. So beginning in the late Archaean there is a rough correspondence between the heat

engine and growth, because crust was added in major episodes that diminished in intensity as time went by.

The growth pulses for individual continents have typically been explained by sweeping many island arcs from a wide oceanic region into one area, and there are well-understood plate-tectonic processes that cause this^{4,5}. The problem is that a large proportion of the Earth's arc-production budget would be needed for many growth pulses. In fact, the worldwide late Archaean pulse would need many times the arc-generating capacity of the modern Earth.

One solution involves upwelling plumes of hot mantle^{6–9}. These can melt, erupt and rapidly put a lot of volcanic material on top of an existing oceanic plate. When such a plateau reaches a subduction zone, it might not be recycled into the mantle, but instead get stuck against the continental margin and become the foundation for a major pulse of crustal growth. This idea was first used to explain early Proterozoic growth¹⁰, but now Stein and Goldstein¹ have applied it to much younger crust in the Arabian–Nubian shield, 900–600 million years (Myr) old.

The authors explain the rapid production of most of the Arabian–Nubian shield using a plume-generated oceanic plateau, formed near the beginning of the activity 900–870 Myr ago. The plateau hit Africa, and a long period of arc-type igneous activity followed, from 850 to 650 Myr ago, reworking and refining the basalt-dominated plateau into the more silica-rich rock typical of continental crust. Indeed, a submarine oceanic plateau may even enhance arc volcanism, because if any of the plateau is subducted, its ocean-altered rock would be expected to increase the flux of water down to magma genesis depths. Additions of arc-type magmas from the subduction of normal oceanic plate presumably occurred too, but the main mass of rock was contributed by the plume-generated plateau. The need to rework basaltic rocks is not unique to this model, as intra-oceanic arcs are also usually basaltic¹¹. Naturally, the reworking of all basaltic materials into more evolved arc-like rock complexes can make it difficult to recognize a plume-related contribution. Stein and Goldstein acknowl-

1. Stein, M. & Goldstein, S. *Nature* **382**, 773–778 (1996).
2. Campbell, I. H. & Taylor, S. R. *Geophys. Res. Lett.* **10**, 1061–1064 (1983).
3. Taylor, S. R. *Nature* **346**, 608–609 (1990).
4. Samson, S. D. & Patchett, P. J. *Aust. J. Earth Sci.* **38**, 595–611 (1991).
5. Sengör, A. M. C. & Natalin, B. A. *Annu. Rev. Earth Planet. Sci.* **24**, 263–337 (1996).
6. Campbell, I. H., Griffiths, R. W. & Hill, R. I. *Nature* **339**, 697–699 (1989).
7. Storey, M., Mahoney, J. J., Kroenke, L. W. & Saunders, A. D. *Geology* **19**, 376–379 (1991).
8. Vervoort, J. D., White, W. M., Thorpe, R. I. & Franklin, J. M. *Econ. Geol.* **88**, 1598–1614 (1993).
9. Kimura, G., Ludden, J. N., Desrochers, J.-P. & Hori, R. *Lithos* **30**, 337–355 (1993).
10. Abouchami, W., Boher, M., Michard, A. & Albarède, F. *J. Geophys. Res.* **95**, 17605–17629 (1990).
11. Rudnick, R. L. *Nature* **378**, 571–578 (1995).
12. Condie, K. C. *Chem. Geol.* **83**, 183–194 (1990).

edge that arc-accretion models are also fully consistent with all the known characteristics of the Arabian–Nubian shield.

That does not obscure the main advantage of a plume-driven crustal genesis model, which is that mantle plumes are a straightforward product of the Earth's heat (from whatever depth they originate), and should have been most common in the Archaean, becoming steadily less so with time. So this is a way of making crust that predicts a closer relationship between the heat engine and continental crust growth.

The intense crustal growth in the late Archaean and early Proterozoic was followed by generally smaller-scale pulses extending to the present day. So a plausible, but by no means established picture is one where a steady trickle of arc volcanism and accretion is added to major pulses of crust growth initiated by accretion of oceanic plateaux. □

P. Jonathan Patchett is in the Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA.

NEUROBIOLOGY

Synapses get smarter

Terrence J. Sejnowski

THE cerebral cortex has long been known to be important for learning and memory, but because of its daunting anatomical complexity it has not been studied as well as the hippocampus, an older and simpler structure. Techniques have improved and it is now possible to pick a pair of cells in a neocortical slice, record from them with whole-cell patch recording, and induce long-term potentiation (LTP) at the synaptic contacts between those cells with Hebbian pairing of presynaptic and postsynaptic action potentials¹.

On page 807 of this issue, Markram and Tsodyks² report that LTP strengthens the response to the first stimulus in a train of stimuli, as expected from hippocampal studies, but can also reduce synaptic responses to subsequent stimuli. This shows that synaptic activity at a cortical excitatory synapse cannot be adequately described by a single number like the strength or 'weight' that is often used in models of neural networks, nor can the effect of LTP be simply described as a strengthening of the synaptic weight. This observation has important consequences for the computational power of synapses in neural networks and the biophysical basis of synaptic plasticity. Synapses may be smarter than we had imagined.

Synaptic strengths are constantly being adjusted from moment to moment. In response to a pair of closely spaced stimuli, the response to the second could get larger or smaller than the first depending on the synapse and its previous history. Short-term facilitation and short-term depression to a train of stimuli occur on time scales ranging from tens of milliseconds to many seconds^{3,4}.

A decade ago, Karl Magleby concluded a review on synaptic plasticity with the prediction that "In addition to long-term changes in synaptic efficacy, which can be assessed with single test pulses, there may be other, more dynamic, long-term changes that can only be detected with patterned stimulation"⁵. Markram and

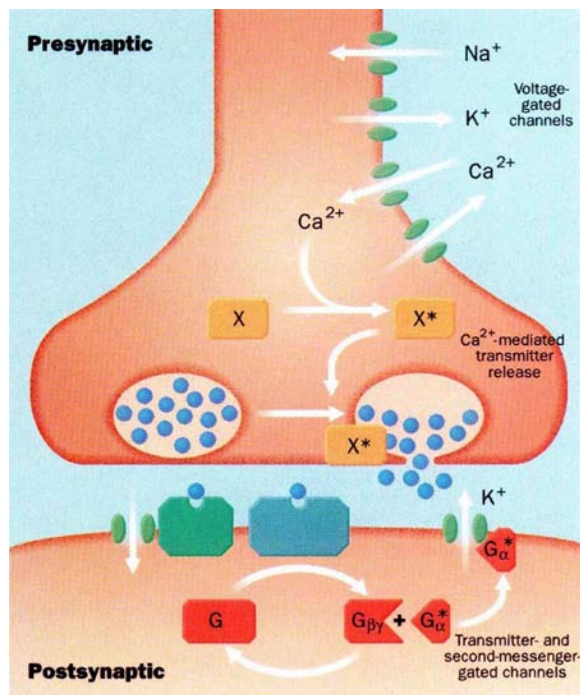
Tsodyks² have now examined long-term changes in short-term synaptic dynamics in neocortical slices and report that the rapidity of short-term depression is substantially and persistently enhanced following LTP.

Although it is possible to develop a phenomenological description of the rates at which short-term synaptic facilitation and depression occur³, we ultimately need to understand the subcellular mechanisms that are involved in synaptic plasticity. A cascade of intracellular events is triggered when calcium enters the presynaptic terminal following invasion by an action potential (see figure)⁶. There is a readily available pool of vesicles that can be mobilized within milliseconds⁷. Short-term depression at peripheral synapses can result from a reduced number of vesicles in the readily releasable pool. The events leading to transmitter release and vesicle recycling can be modelled by kinetic equations, and the rate constants in these equations may be subject to long-term modification⁸. The apparent change in the rate of the short-term depression could result from an increase in the probability of release and subsequent decrease in the population of readily releasable vesicles. These mechanisms make synapses dynamical systems in their own right.

The effects of short-term synaptic dynamics can be counterintuitive. For instance, the steady-state response of a postsynaptic cell to a train of stimuli is independent of firing rate above around 20 Hz (refs 2,

4). This occurs because short-term depression is inversely related to the firing rate, which cancels the term that is proportional to the firing rate. One consequence is that the postsynaptic neuron becomes more sensitive to changes in the firing rate especially when the baseline firing rate is high (L. F. Abbott, personal communication). The changes in short-term dynamics reported by Markram and Tsodyks following LTP would tend to make the cortical network even more sensitive to changes in inputs, because potentiating the response to the first stimulus and more rapidly depressing responses to subsequent stimuli shifts the impact of a train towards the onset of the response (H. Markram, personal communication). The actual impact of LTP *in vivo*, though, could be more complex because the inputs to cortical neurons are highly irregular⁹; when the same irregular train of stimuli is delivered before and after LTP, the responses to some spikes are potentiated, but others can be depressed².

During visual recognition of complex objects, changes in the event-related potential correlated with the category of the object can be registered 150 ms after presentation of a stimulus¹⁰. This is within the range of short-term dynamical changes in strengths of cortical synapses, so synaptic dynamics could be an essential compo-



Synapses are dynamical systems. Presynaptic mechanisms triggered by the entry of calcium into the presynaptic terminal release neurotransmitter molecules into the synaptic cleft, which bind to postsynaptic receptors, and change the conductance of the postsynaptic membrane. In the paper discussed here, Markram and Tsodyks² show that long-term changes in the short-term kinetic parameters governing these steps may lead to changes in the dynamics of synaptic efficacy. Further experiments will be needed to distinguish presynaptic mechanisms from dynamical changes that might also occur at the postsynaptic cell. (Modified from ref. 8.)

nent of object recognition¹¹. The long-term changes now observed in short-term depression, which are synapse specific, may be crucial to understanding the dynamics of memory¹². In neural networks with excitatory recurrent collaterals like those in the cerebral cortex, the positive feedback tends to amplify transient inputs (ref. 13 and C. van Vreeswijk, personal communication). But not all transients will be amplified equally, so only those that match the stored patterns will be filtered and their associations passed onto further stages of processing.

The results reported by Markram and Tsodyks raise various questions which should spur many new studies. What are the effects of LTP on the short-term dynamics at other cortical synapses, particularly those between excitatory and inhibitory cells? How do neuromodulators, which can also change short-term dynamics at synapses, interact with these long-term changes¹³? How does long-term depression affect short-term depression¹⁴? At synapses with low probabilities of release, short-term facilitation can be more prominent than short-term depression¹⁵. How does LTP affect short-term facilitation?

The answers will provide the foundation for approaching many issues in cortical processing, such as how cortical neurons in visual cortex are able to compute the direction of movement of visual stimuli¹⁶, and how novel temporal patterns of inputs are handled¹⁷ — issues that are at the heart of how the cerebral cortex allows us to react intelligently to the changing world. □

Terrence J. Sejnowski is in the Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, California 92037, and the University of California at San Diego, La Jolla, California 92093, USA.

1. Markram, H. & Sakmann, B. *Soc. Neurosci. Abstr.* **21**, 2007 (1995).
2. Markram, H. & Tsodyks, M. *Nature* **382**, 807–810 (1996).
3. Deuchars, J., West, D. C. & Thomson, A. M. *J. Physiol. (Lond.)* **478**, 423–435 (1994).
4. Nelson, S. B., Varela, J. A., Sen, K. & Abbott, L. F. *Soc. Neurosci. Abstr.* **22** (1996). <http://www.faseb.org/sfn96/>
5. Magleby, K. L. in *Synaptic Function* (eds Edelman, G. M. et al.) 21–56 (Wiley, New York, 1987).
6. Sudhof, T. C. *Nature* **375**, 645–653 (1995).
7. Rosenmund, C. & Stevens, C. F. *Neuron* **16**, 1197–1207 (1996).
8. Destexhe, A., Mainen, Z. F. & Sejnowski, T. J. *J. Comput. Neurosci.* **1**, 195–230 (1994).
9. Softky, W. R. & Koch, C. *J. Neurosci.* **13**, 334–350 (1993).
10. Thorpe, S., Fize, D. & Marlot, C. *Nature* **381**, 520–522 (1996).
11. von der Malsburg, C. in *Models of Neural Networks II* (eds Domany, E., van Hemmen, J. L. & Schulten, K.) 95–119 (Springer, Berlin, 1994).
12. Lisberger, S. G. & Sejnowski, T. J. *Nature* **360**, 159–161 (1992).
13. Tsodyks, M. V. & Sejnowski, T. J. *Network* **6**, 111–124 (1995).
14. Bear, M. F. & Malenka, R. C. *Curr. Opin. Neurobiol.* **4**, 389–399 (1994).
15. Stevens, C. F. & Wang, Y. *Neuron* **14**, 795–802 (1995).
16. Abbott, L. F., Sen, K., Varela, J. A. & Nelson, S. B. *Soc. Neurosci. Abstr.* **22** (1996). <http://www.faseb.org/sfn96/>
17. Buonomano, D. V. & Merzenich, M. M. *Science* **267**, 1028–1030 (1995).

Has the fat lady sung?

David Bishop

THIS year is the tenth anniversary of the discovery of high-temperature superconductors. To those of us who have worked on the problem for the past decade, it has been an object lesson in humility. I am reminded of the old joke that you should be careful about what you wish for, as it might come true. If asked ten years ago, most solid-state physicists, including me, might have wished for a liquid-nitrogen-temperature superconductor, believing that it would revolutionize modern electronic technology and society along with it. Just such a class of materials was indeed discovered ten years ago by Bednorz and Müller¹, with superconducting transition temperatures now twice the boiling point of liquid nitrogen. But to the best of my knowledge, a revolution has not yet occurred. Where did we go wrong?

The answer lies partially in the pathological behaviour of the magnetic vortices that penetrate these compounds. How these vortices can misbehave is clearly demonstrated in an experiment described

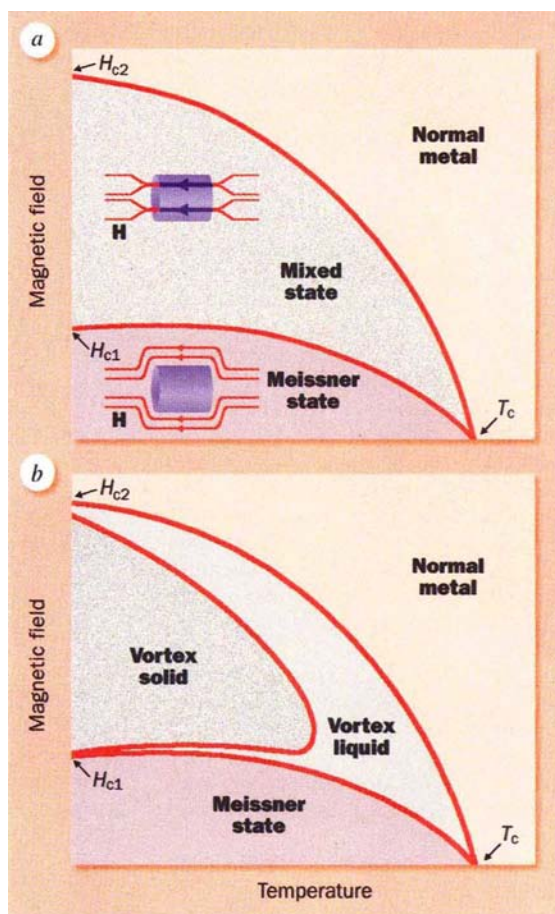
on page 791 of this issue² by Schilling and colleagues. They have confirmed that the magnetic vortex lattice is able to melt into a new state of matter called a vortex liquid. The presence of this liquid means that even modest magnetic fields can cause these materials to be more resistive to the flow of electrical currents than an ordinary piece of copper wire. Therefore, no revolution, or at least one that is going to take a little longer to roll out.

To understand why this happens, one needs to think about how superconductors respond to an applied magnetic field, and about the two classes of superconductor. In general, the superconducting state is inimical to magnetic fields and develops a number of responses to minimize the amount of magnetic flux in its interior. At low fields, below a critical field called H_{c1} , all the flux can be excluded from the interior of a sample — this is the Meissner state. Above a second critical field, H_{c2} , the superconducting state is destroyed and the normal metal is regained.

The first class of superconducting materials, called type I, has $H_{c1}=H_{c2}$. In equilibrium, the field is either totally expelled or superconductivity is completely destroyed. In general, type-I materials are low-field superconductors and are not useful.

All of the technologically important superconductors, including the high- T_c materials, belong to the second class, type II. In these materials $H_{c2}>H_{c1}$, and at intermediate values of the applied field, between H_{c1} and H_{c2} , the field can penetrate but not completely or uniformly (a in the figure). Instead it enters the sample in the form of quantized bundles of magnetic flux called flux lines. The amount of field in each flux line is exactly the same, so the density of lines changes with the applied field. These lines of magnetic flux arrange themselves into a triangular lattice. In the absence of defects this lattice will be a perfectly arranged solid — a solid in the same sense that the chair you are sitting on is a solid: it has a finite shear modulus, whereas liquids and gases have zero shear modulus.

Soon after the high- T_c



a, Phase diagram of a low-temperature type-II superconductor, showing penetration by magnetic vortices in the 'mixed' state. b, High-temperature type-II superconductors have another state, the vortex liquid, which gives them a finite resistance, destroying their usefulness in this region.

superconductors were discovered, it was found that they responded in an unexpected way to an applied magnetic field³. In quite modest fields, the superconducting transitions became smeared out: in some cases the resistivity did not drop to below that of copper until cooled well below the nominal T_c . That observation was the harbinger of difficulties to come.

It is now known that the underlying cause is the pathological behaviour of the magnetic vortices in these systems. In the low-temperature type-II superconductors, the vortex lattice was always found to form a pinned, solid lattice. But in high- T_c materials, it was proposed that thermal fluctuations could cause the ordered vortex lattice to melt (*b* in the figure).

It is easy to see why this is bad for the application of these materials. When a current is applied to a superconductor in the mixed state, it applies a force on each of the magnetic vortices. If the vortices are pinned down so that they cannot move, then no energy is dissipated and a current will flow without loss. If the vortices can move, however, they dissipate some of the energy in the current and the sample then has a finite resistance. The process of pinning the vortex lines is very different for vortex liquids or solids. In the case of a vortex solid, a few pinning centres can hold the entire lattice because it is stiff. Similarly, we can make a carpet stay put with a small number of nails because it is a solid — if it were a liquid, holding it in place would be much more difficult. At low temperatures, the vortex lattice is a solid and the critical current (the maximum current that can be carried without loss) is high. But as the temperature is raised the flux lattice melts and the critical current drops to zero.

This picture of vortex-lattice melting was controversial when it was first suggested^{4,5}, and it has only been generally accepted after years of fighting. The picture of melting is fundamentally a many-body one, in that vortex-vortex interactions dominate the dynamics, forming liquids and solids which then can undergo phase transitions. The competing picture was that the lines are strongly pinned by defects, creating a finite but immeasurably small resistance. In that case there would be no phase transition and the dynamics would be explained by a single-particle picture.

The experiment by Schilling and colleagues² clearly settles the issue. It is well known that phase transitions should always be accompanied by a thermodynamic signature. A first-order (that is, discontinuous) transition should have a latent heat. For vortex lattices this latent heat had never been seen because there are many fewer vortex lines than atoms in a sample, and so the background specific heat is much larger than the signal one wishes to see. Schilling *et al.* have overcome this problem in a most elegant way,

by using a differential calorimeter that allows a precise background subtraction. Using this apparatus, they have clearly seen the latent heat associated with the flux-lattice melting transition in the common high- T_c ceramic superconductor yttrium-barium-copper oxide.

Those of us who are physicists believe that understanding things at the microscopic level is the first step towards solving a problem. We now have a good understanding of how and why flux lattices can melt and what types of matter we are dealing with in the high- T_c superconductors. We can only hope that it will be a short step from understanding to application.

This problem has consumed many of us for almost a decade. In my opinion, the

new experiment ends the discussion. The American baseball player and homespun philosopher Yogi Berra is alleged to have said "it ain't over till the fat lady sings". With regard to the controversy surrounding flux-lattice melting, I think she has. □

David Bishop is at the Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, PO Box 636, Murray Hill, New Jersey 07974-0636, USA.

1. Bednorz, G. J. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
2. Schilling, A. *et al.* *Nature* **382**, 791–793 (1996).
3. Huse, D. A., Fisher, M. P. & Fisher, D. S. *Nature* **368**, 553–559 (1992).
4. Bishop, D. J., Gammel, P. L., Schneemeyer, L. F. & Waszczak, J. V. *Bull. Am. Phys. Soc.* **33**, 606 (1998).
5. Gammel, P. L., Schneemeyer, L. F., Waszczak, J. V. & Bishop, D. J. *Phys. Rev. Lett.* **61**, 1666–1670 (1988).

QUANTUM ELECTRODYNAMICS

Shaking light from the void

Paul Davies

MOST people think that empty space is just that — empty. In fact, even a perfect vacuum teems with activity, as all manner of subatomic particles appear and disappear spontaneously. These 'virtual' particles are the products of quantum processes, and they are permitted a fleeting existence by Heisenberg's uncertainty principle.

The ghostly antics of virtual particles are not merely a theoretical curiosity: they can produce real effects. A famous example is the Casimir effect. Place two mirrors parallel to each other, and a measurable attractive force arises between them. It is caused by the disturbance of the quantum vacuum by the reflecting boundaries. In effect, the mirrors constrain the virtual photons of the vacuum in the cavity between them, and this creates a negative pressure, or tension, which tries to draw the mirrors together.

Twenty years ago, Stephen Fulling and I used a simple mathematical model to investigate the effect of a single mirror on the quantum vacuum (*Proc. R. Soc. Lond. A* **348**, 393–414; 1976 and **356**, 237–257; 1977). From the preliminary results of Bryce DeWitt (*Phys. Rep.* **19C**, 297–357; 1975) we knew that a stationary mirror, or one moving at constant speed, would not disturb the quantum vacuum. Somewhat to our surprise, we found that a uniformly accelerating mirror would also have no effect. If the mirror were to change its rate of acceleration, however, then electromagnetic radiation (real photons) would be produced. The motion of the mirror evidently disturbs the quantum vacuum in a way that excites the electromagnetic

field, effectively promoting virtual photons from the vacuum into real photons.

Although this result was interesting, our calculations indicated that the efficiency of the process is pitifully low. It seemed that only for accelerative changes of unimaginable violence would appreciable numbers of photons be created by the



Intellectual vanity. Einstein admires himself in a mirror, illustrating the invariance of the speed of light. If he were to hit an obstacle, he would also see light boosted from the vacuum by the mirror, accelerating non-uniformly towards him. (From *Einstein for Beginners* by Joseph Schwartz and Michael McGuinness; Icon Books, 1992.)

mirror's motion. This didn't concern us too much at the time, as our main interest in moving-mirror radiation was to provide a heuristic model for the emission of radiation by a black hole — the so-called Hawking effect.

Nevertheless, over the years many imaginative suggestions were made for circumstances where reflecting boundaries might suffer motions of such violence that moving-mirror radiation would be significant. Scenarios included domain wall activity in cosmology, the high-speed collision of atomic nuclei, and the mysterious phenomenon of sonoluminescence, where the passage of sound through water can result in flashes of light being produced.

A new idea for detecting moving-mirror radiation has now been made by Astrid Lambrecht, Marc-Thierry Jaekel and Serge Reynaud (*Phys. Rev. Lett.* **77**,

615–618; 1996). The key feature of their proposal is to use a pair of mirrors in a system similar to that employed in the Casimir effect, but in this case with the mirrors in motion.

The crucial advantage of using two mirrors rather than one is that the space between them can act as a resonant cavity. If the mirror motions consist of simple oscillations with a frequency that matches that of the cavity, then the production of radiation will be hugely amplified. Two proposals are discussed, one in which the cavity length changes sinusoidally, the other where the entire cavity oscillates back and forth rigidly. Both motions can excite a resonant response from the quantum vacuum.

The authors suggest using small, superconducting cavities oscillating through a distance of a few nanometres at GHz frequencies. They estimate that the amplification factor could be as high as one billion, which translates into a production rate of 10 photons per second. Crucially, this production rate occurs with accelerations still gentle enough that the cavity walls will not disintegrate. The created photons trapped within the cavities could be detected with highly excited atoms (sensitive to these low frequencies).

The proposal of Lambrecht *et al.* is the first I am aware of that seeks to produce and detect moving-mirror radiation in a direct way. Their set-up would still create only a tiny quantity of radiation, but confirmation of its existence would provide important support for our theory of the quantum vacuum. A wide range of predictions, from the inflationary universe scenario of the Big Bang to the properties of non-drip paints, depend on this theory.

Moving-mirror radiation also raises important philosophical issues about the nature of space–time and motion. In some respects, the quantum vacuum resembles the old idea of the aether — an invisible medium that was supposed to fill all of space. The quantum vacuum differs from the traditional aether in that it has no privileged reference frame, no state of rest relative to which a material body could be said to move. For that reason, uniform mirror motion has no effect.

More complicated motions, however, do have physical effects. If a moving mirror creates radiation, the energy needed to supply it must be provided by the power source that accelerates the mirror. In turn, the mirror experiences a radiation-reaction force. The fact that a non-uniformly accelerating mirror moving in otherwise empty space experiences a force hints at a deep link between quantum mechanics, relativity theory and cosmology. The nature of that connection is unknown. □

Paul Davies is in the Department of Physics and Mathematical Physics, University of Adelaide, GPO Box 498, Adelaide 5001, Australia.

Retrovirus restriction revealed

John M. Coffin

WITH the identification of the mouse *Fv1* gene, as reported by Best *et al.* on page 826 of this issue¹, one of the biggest fish in retrovirology has at last been pulled from the depths. Starting with studies in the late 1960s and the early 1970s by the late Frank Lilly and colleagues, a number of genes that can prevent disease in mice infected with murine leukaemia virus (MLV) have been found². Some of these genes encode the cell-surface receptor for the virus, with mutations preventing virus entry. Others are MLV-related endogenous proviruses, which are usually defec-

a later step, before, during or after reverse transcription³, and before entry of viral DNA into the nucleus. All retroviruses contain three genes coding for the major groups of virion proteins: *gag* for the internal structural proteins; *pol* for the reverse transcriptase and integrase enzymes; and *env* for the virion surface proteins specifying receptor recognition and entry. Of these, viral genetics unveiled the capsid, or CA, part of the Gag protein as the viral determinant of tropism^{4,5}. This protein forms the core shell of the virion, and is believed to remain in the complex through the metamorphosis of the genome from RNA to DNA and become part of the preintegration complex⁶.

Three features of *Fv1* pique the interest of retrovirologists. First, it acts at some stage of entry, uncoating or transport of the core structure that has been very difficult to access experimentally. Much is known about the mechanism of reverse transcription and integration, much less about their structural and cellular context. Second, natural resistance to infection by a retrovirus might give clues for anti-HIV therapies (*Fv1*-like restriction of HIV infection has been reported at least

once⁷). Finally, *Fv1* is one of the principal remaining puzzles from the early days of retrovirology, and many researchers suffer from unrequited curiosity. Indeed, several of them had cast their lines in the direction of *Fv1*, only to go home empty handed.

The successful strategy employed by Best *et al.* employed both positional cloning and elegant genetics and virology. After preliminary experiments⁸ identified some genetic markers closely linked to *Fv1*, a large cross⁹ was used to identify two markers sufficiently close (about 1 megabase) to select clones from yeast artificial chromosome (YAC) libraries containing the region of interest. The *Fv1*^h-positive YAC was then identified by its ability to confer partial resistance to N-tropic MLV to cells in culture. Eventually, a single open reading frame encoding a 459-amino-acid protein was identified and shown to transfer resistance to N-tropic MLV. A corresponding clone from *Fv1*^{n/n} mice conferred resistance to B-tropic MLV, confirming the identification.

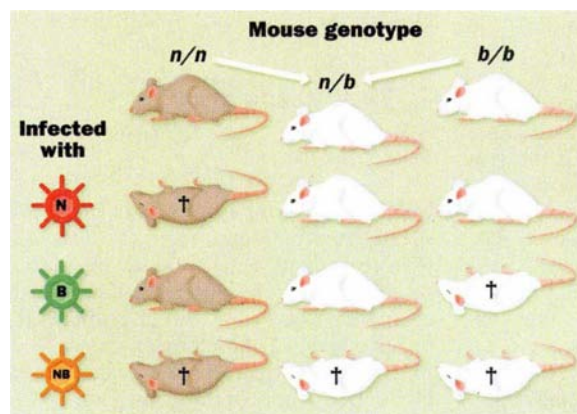


FIG. 1 Genetics of *Fv1*. The two principal alleles, called *n* and *b*, each confer resistance to specific strains of murine leukaemia virus (MLV). Mice carrying the *n* allele resist infection by B-tropic MLV; and mice carrying the *b* allele are resistant to N-type virus. *n/b* heterozygotes are not infectable by either N- or B-type virus, but all three types of animal are sensitive to infection by the third virus type, called NB.

tive but capable of encoding viral-surface (Env) protein which blocks infection by binding to the receptor. Another class of resistance genes does not affect infection, but rather the availability of suitable target cells for disease.

That *Fv1* — named for the Friend (murine leukaemia) virus used to detect it — was different from these was revealed by its unusual genetic properties. Unlike receptor genes, where susceptibility to infection is dominant over resistance, or endogenous proviruses, which display dominant resistance, *Fv1* has multiple alleles and displays a pattern of codominance (Fig. 1) that is most easily explained by postulating that the *Fv1* gene product interacts directly with some virus component to block the replication cycle. Thus the *Fv1*ⁿ product would interact with B-type virus and the *Fv1*^b gene product would interact with N-type virus. NB tropic virus would not interact with either product.

Fv1 does not act at the level of virion entry into cells, contrary to expectations from other resistance genes, but rather at

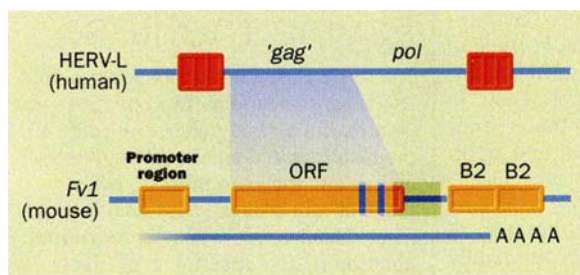


FIG. 2 Structure of the *Fv1* gene. The shaded portions denote regions of difference between the *n* and *b* alleles. The 5' end of the messenger RNA is uncertain; the 3' end lies in sequence derived from the B2 repeat. ORF, open reading frame. The relationship to the human endogenous provirus-like element HERV-L, indicated above, implies that *Fv1* might represent a *gag*-related sequence.

The structure and species distribution of *Fv1* are different from those of a usual gene (Fig. 2). It lacks introns as well as obvious promoter and polyadenylation signals. To date, the promoter has not been identified, but the 3' end has been mapped into a repetitive element (B2) of the coding region. The impression of *Fv1* as a 'neogene', pieced together recently from different parts, is further supported by its distribution. Closely related sequences are found only in laboratory mice and other species of the genus *Mus*. Some species, including cats and rats, show no related sequences at all, and large numbers of distantly related sequences can be seen in other species (including humans). This distribution resembles more closely that of an endogenous provirus than that of a usual gene.

Consistent with this, a search of databases for related sequence turned up a sequence from the *gag* region of a human endogenous provirus-like element (HERV-L), whose *pol* gene resembles that of an unrelated retrovirus, human foamy virus¹⁰. Although this relationship is tantalizing, it should be kept in mind that the sequence lacks detectable relationship to any retroviral *gag* gene as well as hallmarks of the usual *gag* genes. Its identification as a *gag* sequence relies solely on the position of the related HERV-L sequence.

How *Fv1* works should now be accessible to direct experimentation. Even if it is

a *gag* gene, it is unlikely to exert its effect in a simple way. Competition for some cellular factor seems inconsistent with its genetics. The presence in the mouse genome of hundreds of sequences closely related to MLV *gag* genes that exert no obvious suppressive effect argues against simple *gag*-*gag* interaction. Important clues will probably come from comparative analysis of the different alleles, which differ principally by a truncation and substitution of sequence at the carboxy terminus.

A final point of interest regards the evolution of *Fv1* resistance. *Fv1* may represent a rare example of a sequence that originally belonged to a transposable element, but has been separated from it (perhaps by integration of an intact element followed by deletion of sequence on either side) and brought under the influence of transcriptional control elements

from other sources. This process seems to have occurred subsequent to the divergence of *Mus* from other rodents but before the divergence of modern species. Does resistance to MLV infection represent a normal function of this 'gag' sequence? If so, then the ancestral retroelement might represent some sort of antiviral virus, benefiting its host by causing resistance to other viruses. Alternatively, is the resistance a chance effect of mutation of a sequence that happened to be already present when the species was first exposed to MLV? In favour of this possibility is the presence of an apparently non-functional *Fv1* sequence in *Mus dunni*, a species of mouse that shows no sign of prior exposure to MLV infection. Close analysis of the structure and function of the other mouse elements, as well as the human elements, will be highly rewarding. □

John M. Coffin is in the Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA.

NONLINEAR SYSTEMS

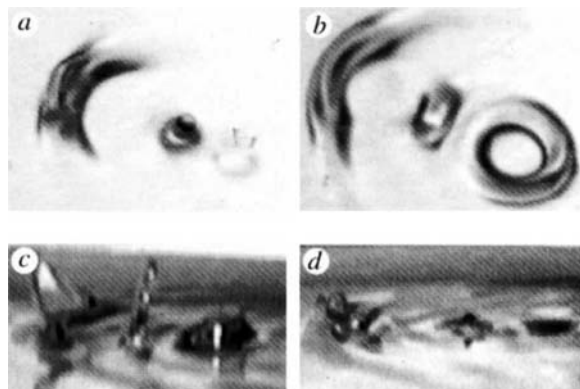
Physics in a jumping sandbox

Jay Fineberg

FROM the stripes of a zebra to rolling sand dunes to the huge cellular structures formed in a thunderstorm, pattern forming systems abound in nature. Although the physical mechanisms that lead to a given type of pattern or structure may be quite diverse, in many cases the formation and subsequent evolution of these states has a universal mathematical description. So the discovery of a new type of structure in a specific system may well be relevant to a large variety of, at first glance, quite different nonlinear systems. On page 793 of this issue¹, Umbanhowar, Melo and Swinney describe the behaviour of a thin, 'sand-like' layer of minute brass balls that are excited into motion by the vertical vibration of their container. At a critical excitation value, strange, well-defined structures form, even though the excitation of the system is spatially uniform.

Previous work by this group revealed that, depending on the strength of the excitation, striped, square or hexagonal structures appear with a characteristic scale many times larger than the grain size.

This large-scale organization is interesting because the actual grains interact with each other simply by inelastic collisions. Now Umbanhowar *et al.* have discovered a surprising and qualitatively different type of structure, which they call an "oscillon". Oscillons are highly localized particle-like excitations of the granular layer which oscillate at half the driving frequency. Once formed, single oscillons are stable. They come in two 'flavours', which like charged particles either repel or attract each other to form dipoles, chains, triangular associations and even lattices.



A typical localized 'solitary' state, propagating from right to left in a highly dissipative fluid³. Two phases relative to the driving frequency are shown. *a* and *b* are views from above, and *c* and *d* are side views of the same states. Compare with Fig. 1 on page 794.

- Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. *Nature* **382**, 826–829 (1996).
- Lilly, F. & Pincus, T. *Adv. Cancer Res.* **17**, 231–277 (1973).
- Pryciak, P. M. & Varmus, H. E. *J. Virol.* **66**, 5959–5966 (1992).
- Rommelaere, J., Donis-Keller, H. & Hopkins, N. *Cell* **16**, 43–50 (1979).
- Ou, C.-Y., Boone, L. R., Koh, C. K., Tennant, R. W. & Yang, W. K. *J. Virol.* **48**, 779–784 (1983).
- Bowerman, B., Brown, P. O., Bishop, J. M. & Varmus, H. E. *Genes Dev.* **3**, 469–478 (1989).
- Williams, L. M. & Cloyd, M. W. *Virology* **184**, 723–728 (1991).
- Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. *J. Virol.* **63**, 1763–1774 (1989).
- Stoye, J. P., Kaushik, N., Jeremiah, S. & Best, S. *Mammal. Genome* **6**, 31–36 (1995).
- Cordonnier, A., Casella, J.-F. & Heidmann, T. *J. Virol.* **69**, 5890–5897 (1995).

Against the grain

Granular media behave in some ways like a solid, in some ways like a liquid or gas, and in some ways like none of those things. Some of their stranger properties are mentioned below. (For a more complete description see ref. 8 and references therein.) Unlike a fluid, sand can heap. Once the slope of a sand pile is increased beyond a critical angle, however, the medium will begin to flow. This flow only occurs within a thin layer near the top of the pile while the remainder of the sand is stationary. To the chagrin of many manufacturers, the same thing happens when trying to force a granular medium down a chute.

Another oddity is that, instead of travelling in a straight line, sound launched into the *side* of a sand layer will curve around and eventually come out of the *top* of the layer. Also, because a granular medium can support weight by the formation of minute 'arches', internal stresses in sand are distributed so that only a small fraction of the grains support the whole pile, and entire regions feel no stress at all. J. F.

A peak-type oscillon appears on the cover of this issue.

This is an example of a driven dissipative nonlinear system having many degrees of freedom, that is, many possible states to choose from. How does such a system typically behave as the level of driving (here, the amount of energy pumped in) is increased? A large class of such systems, as the driving is increased to some threshold value, undergo abrupt transitions called bifurcations, from an initially featureless state to one characterized by a well-defined mode or pattern of collective motion that retains a certain degree of symmetry. As the driving is increased further, this pattern may itself become unstable, undergoing more bifurcations which further reduce the system's symmetry. Dissipation in these systems is not confined to a single region but, like the pattern, is uniformly distributed throughout the medium. Many chemical systems and types of hydrodynamic flow have these characteristics².

But globally distributed patterns are not the only way for nonlinear systems to organize themselves. Highly localized states called solitons were first documented by J. S. Russell in 1834, who observed "a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback... and after a chase of one or two miles, lost it

in the windings of the channel". These intriguing states have since been observed in many diverse nonlinear systems having little or no dissipation. Many types of soliton have the remarkable property that neither their shape nor speed are altered upon collision with other solitons.

Do soliton-like structures exist in natural systems where dissipation plays a major role? As any soldier surrounded by sand-bags can confirm, granular media are highly dissipative. The results of Umbanhowar *et al.* are evidence that localized structures can indeed exist in such systems. How universal are these objects? Soliton-like structures, ubiquitous in non-dissipative systems, have only recently been observed in highly dissipative 2D or 3D systems in experiment³⁻⁶ and theoretical model equations⁷. A fascinating property of both oscillons and the structures observed in references 3 and 7 is that dissipation seems to be *necessary* for their existence. We may, therefore, be looking at a new type of nonlinear object which tends to localize dissipation in driven nonlinear systems.

A surprising aspect of the new experiments is the striking similarity of the observed global patterns in a vibrated granular medium to those observed by replacing the sand with a viscous fluid. The similarity continues with the discovery of oscillons, which are reminiscent of the propagating, localized states recently observed in highly dissipative fluid systems (see figure). Although it is tempting to think of a granular medium as a liquid whose grains behave as fluid 'molecules', granular media (see box above) actually behave in a qualitatively different way from liquids, solids and gases⁸. In fact, despite much active research, no underlying theoretical description for these materials yet exists. The similar behaviour of these two very different types of medium under the same type of excitation may lead to some understanding of the fundamental nature of granular media, as well as of the observed states themselves. □

Jay Fineberg is at the Racah Institute of Physics, The Hebrew University of Jerusalem, Givat Ram 91904, Jerusalem, Israel.

1. Umbanhowar, P. B., Melo, F. & Swinney, H. L. *Nature* **382**, 793-796 (1996).
2. Cross, M. C. & Hohenberg, P. H. *Rev. Mod. Phys.* **65**, 851-1112 (1993).
3. Lioubashevski, O., Arbel, H. & Fineberg, J. *Phys. Rev. Lett.* **76**, 3959-3962 (1996).
4. Steinberg, V., Fineberg, J., Moses, E. & Rehberg, I. *Physica D* **37**, 359-383 (1989).
5. Lerman, K., Bodenschatz, E., Cannell, D. S. & Ahlers, G. *Phys. Rev. Lett.* **70**, 3572-3575 (1993).
6. Rotermund, H. H., Jakobith, S., von Oertzen, A. & Ertl, G. *Phys. Rev. Lett.* **66**, 3083-3086 (1991).
7. Deissler, R. J. & Brand, H. R. *Phys. Rev. Lett.* **74**, 4847-4850 (1995).
8. Jaeger, H. M. & Nagel, S. R. *Science* **255**, 1523-1531 (1992).

DAEDALUS

Laying the lies

THE Anglo-Saxon system of adversarial law is very imperfect. Two opposite sides — prosecution and defence — trade all the distorted and emotional arguments and insults they can think of, while a judge or jury tries to guess who is lying most. The procedure shares the drama, philosophy, and uncertainty of tribal warfare or big-league spectator sport. Nobody seriously looking for the truth (as in a scientific investigation or an air crash enquiry) would dream of using it.

The crucial problem, says Daedalus, is the ease with which human beings can tell lies. Indeed, this is almost a defining human skill. Many animals can lie in a simplistic way — pet cats, for example, will heroically pretend not to have been fed. But only human beings are self-conscious enough to elevate lying into a way of life, a vital skill in the complex competitive social game. One theory even claims that the human unconscious mind has evolved specifically to hold the secret truth, so that the conscious mind can lie boldly without stumbling over contradictions. Hence the role of the subconscious in psychology.

Now modern brain-scanning techniques can locate the active site of the brain at any moment. Daedalus reckons that telling the truth should activate just one site, where the relevant information is stored. Telling a lie should activate two sites, one holding the lie and the other holding the concealed truth. Positron-emission tomography is perhaps the best locator, but its radioactivity is worrying. So Daedalus plans to use magnetic resonance. DREADCO physicists are now injecting ¹³C glucose into the blood of volunteers ranging from born-again Quakers to car salesmen, firing at them questions ranging from innocent to highly incriminating, and using ¹³C NMR tomography to locate the releases of ¹³CO₂ evoked in their brains. The results should reveal the brain-signature of lying. With luck, it will be simple, consistent, and easily distinguished from the effects of emotional stress.

Daedalus will then devise an NMR witness box, or even helmet, to detect and display this signature. It will transform our legal system. Guilty criminals will no longer plead innocent, and lying plaintiffs will no longer risk litigation. False claims of liability, negligence, harassment and abuse will swiftly fade away. NMR helmets for the lawyers themselves would be even more salutary. Many uses for the system beckon outside the court-room, too; but Daedalus is wary. Such a social nuclear deterrent is best kept in reserve as a threat. If widely deployed, it could make social life quite impossible. David Jones

Climate and species' range

SIR — Though physical evidence of global warming continues to accumulate¹⁻³, it is less clear whether the predicted biological consequences are occurring. The key prediction is that species' ranges should move both polewards in latitude and upwards in elevation⁴⁻⁷. To date, published claims of species' range shifts have extrapolated regional patterns from observations on a small part of a species' range. These studies all assume that changes in population density at a point⁸⁻¹¹ or movement of a boundary at one latitudinal end of a range^{12,13} can be unambiguously interpreted as range shifts, rather than as merely local density changes, range expansions or contractions. Reliable evidence for range shifts must include examination of a species' entire range. I report here the first study to provide evidence of the predicted range shifts.

I censused populations of Edith's checkerspot butterfly (*Euphydryas editha*) throughout its range, and found significant latitudinal and altitudinal clines in

population extinctions at sites undegraded by human activities, producing a northwards and upwards shift in the species' range. I documented extinction and persistence in 151 previously recorded populations of this butterfly, and excluded from the data set all sites where butterfly habitat was degraded and no longer usable by this species, including sites altered by human activities such as land-clearing, construction, overgrazing and introduction of exotic plants. Dense populations of this butterfly have persisted even in moderately disturbed (grazed or logged) sites, if enough suitable host plants remain.

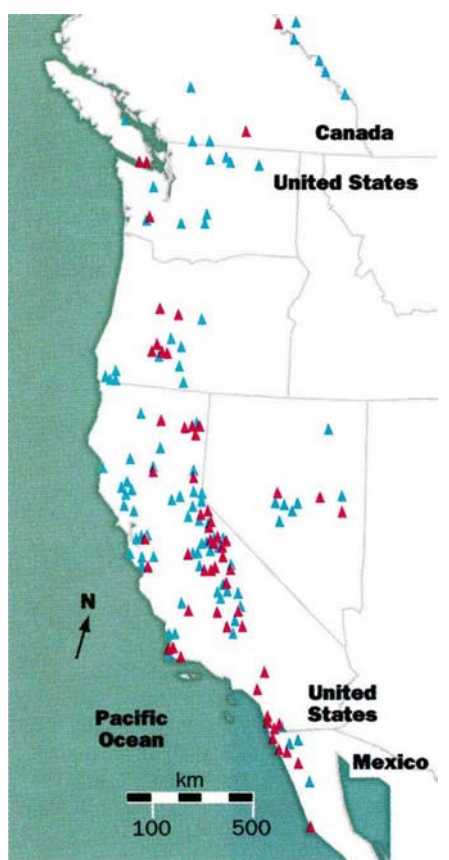
Edith's checkerspot occurs as a mosaic of sedentary, discrete populations with intermittent colonizations, extinctions and re-colonizations¹⁴⁻¹⁷. At equilibrium, the rate of extinction should be constant and equal to the rate of colonization and recolonization. Climate warming should cause net extinctions to increase in the south and at low elevations and to decrease in the north and at high eleva-

tions. It is the net extinction rate that is of interest in the context of global warming.

I found a striking latitudinal cline in net extinction rates (Figs 1, 2a; $P = 0.009$). Sites where previously recorded populations still existed were on average 2° further north than sites where populations were extinct. Populations in Mexico were four times more likely to be extinct than those in Canada. Net extinctions also significantly decreased with altitude (Figs 1, 2b, $P = 0.04$). Populations above 2,400 m were significantly more persistent than those at all lower elevations ($P = 0.016$). Although a predicted result of climate warming is an increased extinction rate at the very lowest elevations, no such trend appears in the data.

Any factor affecting recolonization might vary systematically and produce the observed clines in net extinctions. However, recolonization was rare over long time periods and over a large portion of the range. Among 31 populations that were repeatedly visited after a recorded extinction, a maximum of 13% were recolonized over a 30-year period. Further, the degree of isolation of populations (which would directly affect recolonizations) was historically similar at the extremes of the range. Over the past century, records documented 0.32 populations per 1,000 km² in mainland Canada and 0.87 populations per 1,000 km² in Mexico (density estimates restricted to the species range in each area). Recent human destruction of butterfly habitat and extirpation of populations could, by increasing isolation, reduce recolonization rates and inflate net extinction rates even in undegraded habitats. However, habitat degradation was latitudi-

FIG. 1 Map of censused sites. Extinct populations are shown by red triangles, present populations by blue triangles. Because of space limitations, three censused populations in Colorado are not shown but are available from the author on request. I compiled historical population records from museum specimens, private collections and researchers' field notes. Between 1992 and 1996, I censused 115 sites with historical records to classify their current status as extinct or intact, and for 36 additional sites determined current status using information provided by K. Agnew, G. Austin, D. Bauer, D. A. Boughton, J. Brown, P. Chai, P. R. Ehrlich, J. Emmel, G. Gilchrist, C. Guppy, S. P. Harrison, D. Kinsinger, N. Kondla, A. Launer, S. O. Mattoon, A. R. Moldenke, D. D. Murphy, O. Shields, J. Shepherd, M. C. Singer, W. L. Swisher, C. D. Thomas, D. Vasco and R. R. White. I controlled for time elapsed since the initial record by sampling sites within each geographic region that represented the range of historically recorded dates. I did not census sites that: (1) could not be unambiguously located from the records; or (2) were degraded by loss of usable host plants. I censused all butterfly life stages at each included site, and considered sites less than 5 km apart to be part of the same population¹⁷. I visited sites during the flight season and searched for adults in suitable habitat within the general vicinity of the historical record (resulting in areas of 0.25 km² to 24 km² searched). If I found no adults, I searched potential host plants for egg clusters and larval webs, checking 100% of the host plants if the habitat was small. In larger habitats, I searched approximately 500 host plants or 50% of the plants present, whichever was smaller. More than 50% of the time that I found populations, I discovered an egg mass or larval web within the first 30 plants.



- Oechel, W. C. & Vourlitis, G. L. *Trends Ecol. Evol.* **9**, 324-329 (1994).
- Kerr, R. A. *Science* **270**, 1565-1566 (1995).
- Thomason, D. J. *Science* **268**, 59-68 (1995).
- Peters, R. L. & Darling, J. D. *Bioscience* **35**, 707 (1985).
- Woodward, F. I. in *Global Warming and Biological Diversity* (eds Peters, R. L. & Lovejoy, T. E.) 105-123 (Yale Univ. Press, 1992).
- Davis, M. B. & Zabinski, C. in *Global Warming and Biological Diversity* (eds Peters, R. L. & Lovejoy, T. E.) 297-308 (Yale Univ. Press, 1992).
- Schneider, S. H. in *Biotic Interactions and Global Change* (eds Kareiva, P. et al.) 9-23 (Sinauer, Sunderland, Massachusetts, 1993).
- Southward, A. J., Hawkins, S. J. & Burrows, M. T. J. *Therm. Biol.* **20**, 127-155 (1995).
- Barry, J. P. et al. *Science* **267**, 672-675 (1995).
- Chapin, F. S. III *Ecology* **76**, 694-711 (1995).
- Roemich, D. & McGowan J. *Science* **267**, 1324-1326 (1995).
- Kullman, L. *Can. J. Forest. Res.* **16**, 761-771 (1986).
- Pollard, E. & Eversham, B. C. in *Ecology and Conservation of Butterflies* (ed. Pullin, A. S.) 23-36 (Chapman & Hall, London, 1995).
- Ehrlich, P. R. et al. *Oecologia* **46**, 101-105 (1980).
- Ehrlich, P. R. *Science* **134**, 108-109 (1961).
- Ehrlich, P. R. *Evolution* **19**, 327-336 (1965).
- Harrison, S., Murphy, D. D. & Ehrlich, P. R. *Am. Nat.* **132**, 360-382 (1988).
- Singer, M. C. & Ehrlich, P. R. *Fortschritte der Zoologie* **25**, 53-60 (1979).
- Murphy, D. D. & Ehrlich, P. R. *J. Lepid. Soc.* **34**, 316-320 (1980).
- Thomas, C. D., Singer, M. C. & Boughton, D. *Am. Nat.* (in the press).
- Singer, M. C. & Thomas, C. D. *Am. Nat.* (in the press).

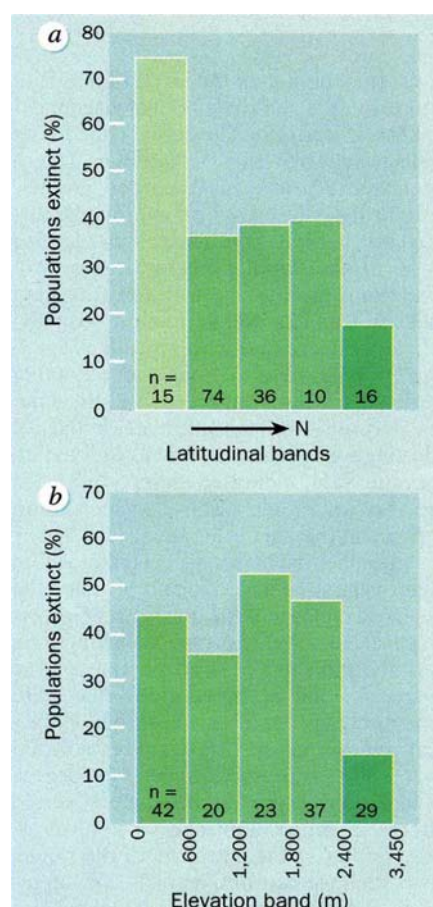


FIG. 2 Proportions of extinct populations in five latitudinal (a) and elevational (b) bands. a, Analysis with latitude as a continuous variable: Mann-Whitney rank test comparing latitudes of the two state groups, extinct and persistent (mean extinct group = 38.5°N, mean present group = 40.6°N, tied Z-value = -2.60, $n=151$, $P=0.009$). Analysis with latitude as a categorical variable: to test for significant break points, I analysed extinctions by latitudinal bands with a 5×2 contingency table (5 levels of latitude evenly divided from 30° N to 53° N; 2 levels of status, extinct or present) using a log-likelihood ratio test: $G=15.75$, d.f. = 4, $P=0.003$. I performed post-hoc sub-divided analyses to test for significant differences among adjacent bands. Latitudinal bands of different shades of green are significantly different from each other at $P \leq 0.05$. b, Analysis with altitude as a continuous variable: Mann-Whitney rank test on altitude between the two state groups, extinct and persistent (mean extinct group = 1,280 m, mean present group = 1,585 m, tied Z-value = -2.05, $n=151$, $P=0.04$). To test the significance of an apparent break point at 2,400 m, I analysed extinctions by elevational bands with a 5×2 contingency table (5 levels of elevation; 2 levels of status, extinct or present) using a log-likelihood ratio test: $G=12.16$, d.f. = 4, $P=0.016$. I performed post-hoc sub-divided analyses to test for significant differences among subsets of the elevational bands. Elevational bands of different shades of green are significantly different from each other at $P \leq 0.05$.

nally symmetrical: low at the extremes of the range, with about 20 % of previously recorded sites degraded in Mexico ($n=10$) and 17 % in mainland Canada ($n=24$), and higher for all other latitudes.

Thus, it is unlikely that the observed latitudinal cline in net extinctions was caused by differences in initial population isolation or subsequent land-use changes. This result, in conjunction with earlier detailed studies of climate-caused population extinctions in this butterfly^{14,18-21}, suggests climate change as the cause of the observed range shift. However, conclusive evidence for or against the existence of the predicted biological effects of climate

change will come, not from attempts to analyse all possible confounding variables in single studies such as this one, but from replication of this type of study with additional taxa in other regions. Until this has been done, the evidence presented here provides the clearest indication to date that global climate warming is already influencing species' distributions.

Camille Parmesan

National Center for Ecological Analysis and Synthesis,
735 State Street, Suite 300,
University of California at Santa Barbara,
Santa Barbara, California 93101-3351, USA
e-mail: parmesan@nceas.ucsb.edu

Base-rate errors and rain forecasts

SIR — The nature of the base-rate error — the neglect of prior probabilities in judging the probability of events — has recently been discussed¹. Yet despite its potentially serious implications for many real-life issues, the base-rate error has yet to achieve wider recognition. I would therefore like to draw to readers' attention the effect of the base-rate error on a familiar (indeed, notorious) dilemma — that of how to respond to weather forecasts.

It seems obvious that decisions affected by the weather (going for a walk, for example) are best made by putting one's faith in the most accurate forecast available. Surprisingly, however, the base-rate

effect can make this a sub-optimal approach.

The UK Meteorological Office's 24-hour forecasts of rain currently achieve around 83 per cent accuracy, while the probability of rain on the hourly timescale relevant to walks is around 0.08. The table reveals the impact of the

THE VARIOUS OUTCOMES OF FORECAST AND WEATHER OVER 1,000 1-HOUR WALKS

	Rain	No rain	Sum
Forecast of rain	66	156	222
Forecast of no rain	14	764	778
Sum	80	920	1,000

base-rate error in the interpretation of forecasts of rain. With forecast accuracies of 83 per cent, one might expect that a forecast of rain during the one hour walk would be correct 83 per cent of the time. However, the hourly base-rate of rain in the United Kingdom is so low that forecasts of rain are more than twice as likely to be wrong as right: from the table, the probability of rain, given a forecast of rain — that is, $P(\text{rain}|\text{forecast of rain})$ — is $66/222=0.30$, whereas $P(\text{no rain}|\text{forecast})=156/222=0.70$.

This result suggests that those who ignore Meteorological Office forecasts may fare better than those who abide by them. A decision-theoretic analysis shows that this is indeed the case. Let R, F and W represent the events of rain falling during the walk, rain being forecast, and going on the walk, respectively. Then

$$P(R \& W) = P(R)[P(F|R).P(W|F) + P(\sim F|R).P(W|\sim F)] \quad (1)$$

with similar expressions for the three other permutations of R, W and negations $\sim R$, $\sim W$. The forecasting accuracy is captured by $P(F|R) = A$ and $P(F|\sim R) = B$, while the responses to the forecasts are represented by $P(W|F) = m$ and $P(W|\sim F) = n$; (1) then becomes

$$P(R \& W) = P(R)[Am + (1-A)n] \text{ etc} \quad (2)$$

Let L_{tot} represent the loss function, comprising the losses resulting from the outcomes of the various decisions:

$$L_{\text{tot}} = L_{00}P(R \& W) + L_{11}P(\sim R \& \sim W) + L_{10}P(\sim R \& W) + L_{01}P(R \& \sim W) \quad (3)$$

Optimal strategies minimize L_{tot} . Substituting from equation (2), and keeping only terms in m and n ,

$$L_{\text{tot}} \sim m\{P(R)AK - P(\sim R)B\} + n\{P(R)(1-A)K - P(\sim R)(1-B)\} \quad (4)$$

$$\text{where } K = (L_{00} - L_{01}) / (L_{11} - L_{10}) \quad (5)$$

represents the relative losses resulting from the outcomes in equation (3). With $A=0.83$, $B=0.17$, $P(R)=0.08$, we find that basing our decision on Meteorological Office forecasts ($m=0$, $n=1$) gives $L_{\text{tot}} \sim (K-56)/74$, whereas ignoring forecasts of rain ($m=n=1$) gives $L_{\text{tot}} \sim (K-12)/13$. Thus unless one is particularly concerned about getting wet ($K > 2$), the base-rate effect makes disregard of forecasts of rain the optimal strategy.

Similar reasoning also reveals that, contrary to popular belief, always carrying an umbrella is a sub-optimal strategy unless one is morbidly afraid of getting wet ($K > 56$). Indeed, unless $K > 12$, the base-rate effect makes even insouciant optimism a better strategy.

Robert A. J. Matthews

Department of Computer Science,
Aston University,
Birmingham B4 7ET, UK

1. Goodie, A. S. & Fantino, E. *Nature* **380**, 247–249 (1996).

Chemokines and HIV replication

SIR — Three chemotactic peptides belonging to the β -chemokine group of cytokines, RANTES, MIP-1 α , and MIP-1 β , inhibit the infection of CD4⁺ T cells by primary HIV-1 strains¹. Although all three peptides suppress HIV-1 replication in T cells, they do so with different efficiency, with RANTES being the most active, followed by MIP-1 β , then MIP-1 α ¹. Here we demonstrate that, in sharp contrast to their observed antiviral effects in T cells, β -chemokines actually stimulate the replication of primary HIV-1 strains in macrophages, another major target of this virus. The magnitude of stimulation is dependent on the cell donor and HIV-1 strain used, and varies for different β -chemokine peptides.

In addition to T lymphocytes, monocyte/macrophages are a rich source of β -chemokines in the body². HIV-1 infection itself upregulates β -chemokine expression in monocytes both *in vitro* and *in vivo*³. In the light of their anti-HIV effects on T cells, the release of these peptides by macrophages in response to HIV-1 infection might reflect a defensive manoeuvre of the immune system. To test this hypothesis, we performed experiments to assess the anti-HIV-1 activity of β -chemokine peptides in monocyte cultures, and the effect of chemokine peptides on HIV-1 replication in T lymphocytes.

We used two primary HIV-1 isolates that replicate both in T lymphocytes and monocytes, 92US657 and 92US660 (obtained from NIH AIDS Research and Reference Reagent Program), for infections. We isolated lymphocytes from the same donors as monocytes by passing non-adherent cells through a T-cell enrichment column (R&D, Minneapolis) resulting in a 95% pure CD4⁺ + CD8⁺ T-cell population. As expected, RANTES,

MIP-1 β and MIP-1 α (all peptides from PeproTech Inc.) suppress replication of both isolates in T lymphocytes, with a 50% inhibitory concentration (IC₅₀) of 5, 15 and 35 ng ml⁻¹, respectively, in good agreement with previously published results¹. At 500 ng ml⁻¹, each chemokine inhibited HIV-1 replication by more than 95 %, and this concentration was chosen for testing chemokines' effects on HIV-1 infection of macrophages.

In contrast, we observed an enhancing, rather than an inhibitory, activity of all three β -chemokines on the replication of both HIV-1 strains in macrophage cultures prepared from two donors (see figure). We observed this enhancing effect of β -chemokine peptides throughout the course of infection; data in the figure show results obtained on day 15 post-infection when virus replication reached the peak. This effect is dose-dependent (data not shown) and is not the result of contaminating lipopolysaccharide in chemokine preparations (less than 0.1 ng per μ g of peptide) as similar amounts of lipopolysaccharide either have no effect or suppress HIV-1 replication in macrophage cultures (not shown). We obtained similar results (not shown) with another HIV-1 strain, HIV-1_{ADA}, and β -chemokine peptides from a different source (R&D), indicating that the observed phenomenon does not result from peculiarities of the chemokine preparations but rather reflects a general feature of these molecules.

Dragic *et al.*⁴ have also observed failure of β -chemokines to inhibit HIV-1 infection of primary monocytes. However, these authors did not detect any stimulatory effect of these peptides on viral replication. The different results could be due to the source of virus used for infection:

Dragic *et al.* used recombinant HIV-1 strains, whereas our results were obtained with primary isolates. As shown in the figure, strain differences significantly affect the magnitude of the stimulation. Although the observed stimulatory effect of β -chemokines on HIV-1 replication could be a by-product of their ability to activate macrophages, other mechanisms must be also involved, as MIP-1 α is a much more potent stimulatory agent than MIP-1 β ⁵, yet their effect on HIV-1 replication is comparable in most cases (a in the figure).

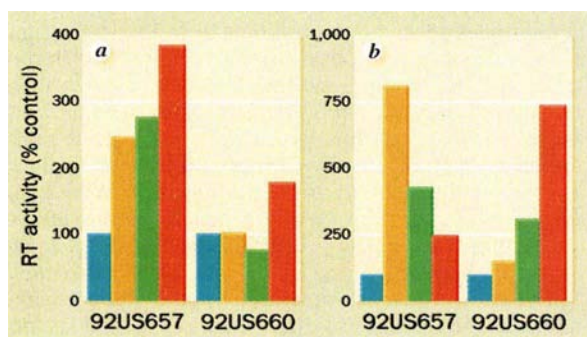
Taken together, these findings suggest that the ability of β -chemokine peptides to either inhibit or stimulate HIV-1 replication is cell-type dependent. Failure of chemokines to inhibit HIV-1 replication in macrophages despite the presence on these cells of CCR-5, a β -chemokine receptor and a co-receptor for HIV-1^{4,6}, indicates that a different co-receptor which does not bind β -chemokines might be involved in HIV-1 infection of macrophages. Because HIV-1 infection induces β -chemokine expression in monocytes³, the infected immune system might actually benefit by allowing tempered viral replication in these cells (which are less susceptible to the cytopathic effects of infection than T cells⁷), so that sufficient amounts of the β -chemokines can be produced to inhibit vigorous virus replication in T-lymphocyte populations. On the other hand, high levels of β -chemokine peptides could produce harmful results by enhancing HIV-1 replication in macrophages and/or intensifying virus-induced inflammation, as demonstrated for Cocksackie and influenza viruses⁸. These considerations should be taken into account when considering the use of β -chemokines as anti-HIV therapeutic agents.

Helena Schmidtayerova*, Barbara Sherry & Michael Bukrinsky
The Picower Institute for Medical Research,
Manhasset, New York 11030, USA
e-mail: mbukrinsky@picower.edu

* Also at: Institute of Virology, Slovak Academy of Sciences, 842 46 Bratislava, Slovakia.

1. Cocchi, F. *et al.* *Science* **270**, 1811–1816 (1995).
2. Sherry, B. & Cerami, A. *Curr. Opin. Immunol.* **3**, 56–60 (1991).
3. Schmidtayerova, H. *et al.* *Proc. Natl. Acad. Sci. USA* **93**, 700–704 (1996).
4. Dragic, T. *et al.* *Nature* **381**, 667–673 (1996).
5. Fahey, T. J. *et al.* *J. Immunol.* **148**, 2764–2769 (1992).
6. Deng, H. *et al.* *Nature* **381**, 661–666 (1996).
7. Gendelman, H. E. *et al.* *AIDS* **3**, 475–495 (1989).
8. Cook, D. N. *et al.* *Science* **269**, 1583–1585 (1995).

Effect of β -chemokine peptides on HIV-1 replication in monocyte cultures. Dark boxes, controls; light hatching, MIP-1 α ; heavy hatching, MIP-1 β ; reverse hatching, RANTES. We cultured macrophage cultures prepared from PBMCs of two donors (a and b) by adherence to plastic³ for 7 days and then infected them with HIV-1 (2×10^4 c.p.m. of reverse transcriptase (RT) per 10^6 cells in 1 ml medium) in the presence of 500 ng ml⁻¹ chemokine β -peptides. We replaced the culture medium every 3 days with fresh medium containing 500 ng ml⁻¹ chemokines. Fifteen days after infection, we tested culture supernatants for reverse transcriptase activity. Results are presented as per cent of RT activity in untreated (control) culture supernatants taken as 100%. RT activity for the controls is 5.8×10^5 c.p.m. ml⁻¹ for 92US657 and 2.8×10^5 for 92US660. Blue, control; orange, MIP-1 α ; green, MIP-1 β ; red, RANTES.



Scientific Correspondence

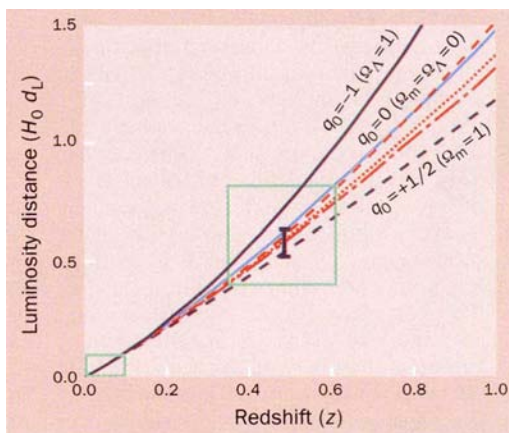
Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.

Missing energy and cosmic expansion

SIR—Evidence is mounting that the ordinary (baryonic) matter and the dark matter in the Universe sum to less than the critical density required for a cosmologically flat Universe¹. Either space is curved (the Universe is open) or there is some additional “missing energy” in the Universe. Inflationary cosmology predicts a flat Universe and, hence, favours

standard candles whose luminosity is inversely proportional to the square of their distance. (This allows their use to determine to a fairly high accuracy the absolute distances to the galaxies in which the supernovae occur.) To date, the strategy has been to use measurements at low redshift $z \leq 0.1$ to determine precisely the linear (Hubble law)

Demonstration that the luminosity distance versus redshift prediction depends on both q_0 and α of the missing energy. All curves correspond to a common value of H_0 , as measured by observations at small redshift (lower green dashed box). Attempts to measure q_0 entail measurements at moderate redshift (upper green dashed box). Labelled, dashed curves, canonical limiting cases of $q_0 = -1$, 0 or $+1/2$, assuming a universe with only vacuum density, spatial curvature or matter energy, respectively. Red curves correspond to $q_0 = 0$, but diverge from one another because the missing energy obeys a different equation of state: dotted curve corresponds to $\Omega_m = 1/3$ and $\Omega_u = 2/3$ and $\alpha = -1/2$; the dot-dashed curve corresponds to $\Omega_m = 2/3$, $\Omega_u = 1/3$ (Ω_u corresponds to missing energy with $\alpha = -1$). Note also that the canonical $q_0 = 0$ curve (red, dashed) is nearly degenerate with the blue solid curve even though the latter has $q_0 = -0.4$ and a substantial cosmological constant, $\Omega_\Lambda = 0.6$. The error bar represents roughly the current uncertainty from supernovae measurements; reduced uncertainty, by a factor of five or more, may be possible in near-future, systematic searches.



the missing energy explanation. (Because inflation is the model currently favoured by theorists, I will be concerned only with that model for now.) For historical reasons, the missing energy is often assumed to be the energy density associated with the vacuum state of the Universe; this was first introduced by Einstein in terms of a cosmological constant (Λ). The energy density of the vacuum remains unchanged as the Universe expands. However, it is important to realize that other forms of missing energy are possible. For example, the energy density could be due to interacting fields (a scalar field rolling down a potential) or topological defects (such as cosmic strings) whose energy density changes as the Universe expands. Here I show that the form of the missing energy affects the interpretation of observations already under way.

Recent attempts to measure the deceleration of the expanding Universe using the observed redshifts z of luminous sources out to $z \approx 0.5$ have tacitly assumed a form for the missing energy. Removing the assumption leads to a reinterpretation of current results and future prospects.

This consideration is timely in that systematic efforts are under way to measure the magnitudes and redshifts of distant type Ia supernovae, which appear to be

distance–redshift relation², and then focus on measuring the deviation from that linear law^{3,4} using supernovae at moderate redshifts $z = 0.35$ – 0.6 . The deviation from the linear law has been conventionally interpreted as depending only on the present deceleration rate q_0 . This is not correct.

The deviation from the linear law also depends on the equation-of-state α , defined as the ratio of pressure to energy density of the missing energy. If the evidence for subcritical matter energy density continues to grow, determining α will emerge as a major observational challenge for cosmology. As the Universe expands, the missing energy density varies as $(\text{volume})^{-(1+\alpha)}$. Ordinary or dark (pressureless) matter energy density falls inversely with volume. For the traditional cosmological constant, $\alpha = -1$, but for interacting fields and topological defects α can vary between -1 and 0 .

The issue is best demonstrated by considering the “luminosity distance–redshift relation”, the relation determined by the observations of supernovae. The luminosity distance between a given source and us is defined as $d_L^2 \equiv L/4\pi F$ where L is the emitted energy per unit time and F is the energy received per unit time. An elementary calculation shows that $d_L = (1+z)r_1$ where the comoving distance r_1 satisfies:

$$\int_0^{r_1} \frac{dr}{(1-kr^2)^{1/2}} = \int_0^z \frac{dz' [H_0(1+z')]^{-1}}{[\Omega_m(1-z') + \Omega_u(1+z')^{1+3\alpha} + (1-\Omega_m-\Omega_u)]^{1/2}} \quad (1)$$

here H_0 is the Hubble constant, and Ω_m and Ω_u (Ω_{unknown}) are the ratios of the matter density and the missing energy density, respectively, to the critical density (the total energy density in a flat universe). Expanding equation (1) for small z , the luminosity distance–redshift relation is

$$H_0 d_L = z + \frac{1}{2}(1-q_0)z^2 + \mathcal{O}(z^3)$$

where

$$q_0 = \frac{1}{2} \Omega_m + \left(\frac{1+3\alpha}{2}\right) \Omega_u$$

To lowest order, the deviation from the linear Hubble law depends on q_0 alone. However, at moderate redshift ($z = 0.35$ – 0.6), the higher-order $\mathcal{O}(z^2)$ contributions to $H_0 d_L$ are non-negligible, and these depend on the equation of state of the missing energy. The figure illustrates the point. It shows how the luminosity distance–redshift relation differs for two models with the same q_0 and H_0 . It also shows how two models with very different values of q_0 and α lead to nearly degenerate predictions. The dependence on α has been omitted in conventional treatments^{3,5}.

Hence, the current strategy of measuring the deviation from the linear Hubble law over a narrow range of redshift cannot determine q_0 or Ω_m without specific assumptions about the nature of missing energy. In particular, current limits do not exclude a substantial cosmological constant. Expanding the observational strategy to the range from $z = 0.1$ to 1 can provide a simultaneous tight constraint on q_0 and α . Even here, there can remain near degeneracies, as in the figure, but typically between starkly different models that can be easily differentiated by other observational constraints (for example, lower bounds on the matter density). By this combined approach, it remains possible to determine q_0 and the equation of state of the missing energy of the Universe.

Paul J. Steinhardt

Department of Physics and Astronomy,
University of Pennsylvania,
Philadelphia, Pennsylvania 19104, USA

- Ostriker, J. & Steinhardt, P. J. *Nature* **377**, 600–603 (1995).
- Riess, A. G., Press, W. H. & Kirshner, R. P. *Astrophys. J.* **438**, L17–L20 (1995).
- Perlmutter, S. et al. *Astrophys. J.* **440**, L41–L44 (1995).
- Perlmutter, S. et al. *Preprint astro-ph/9602122* (1996).
- Riess, A. G., Stabell, R. & de Lange, F. G. *Mem. R. Astron. Soc.* **71**, 143–248 (1967).
- Goobar, A. & Perlmutter, S. *Astrophys. J.* **450**, 14–18 (1995).

Lighter than air

John L. Casti

The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age. By John Horgan. *Helix* (Addison-Wesley): 1996. Pp. 309. \$24.

QUESTIONS about origins have always held a strong fascination for the intellectually inclined, perhaps because such one-time-only events are difficult to study, so providing a vast playground for unbridled speculation and almost limitless armchair philosophy. Equally fascinating, it seems, are the no-time-only events of how things end. Several Cassandras have publicly aired their angst over the incipient demise of something beloved in recent years, ranging from Steven Weinberg's dreams of a final theory in particle physics to Francis Fukuyama's musings on the end of history. The latest addition to this cast of doom-sayers is the *Scientific American* journalist John Horgan, who ups the ante by trumpeting the imminent end of *all* science. What can one possibly make of such a temerarious claim?

Horgan states that "scientists have answered some questions beyond reasonable doubt, and have stitched these findings into a compelling, if not terribly detailed, map and history of existence". In other words, once one has discovered the Standard Model of particle physics and knows the double-helix structure of DNA and that the Universe is expanding, what is left? Putting these pieces of knowledge together with the threats to science from technophobes, religious fundamentalists, deconstructionist literary critics and animal-rights advocates, not to mention penny-pinching politicians, who can doubt that the scientific enterprise is grinding to a halt? Or so goes Horgan's argument.

Science as a process

Science is a process of a very special type for getting at the scheme of things. What distinguishes it from religion, mysticism, poetry and all the other players in the reality-generation game is that it provides answers to questions about the world by invoking a set of rules (that is, a theory, formula or program). But not just any old rule will do. A scientific rule has certain properties — public accessibility, clarity, brevity, lack of bias — and is generated by following a definite procedure, the so-called scientific method. So if science is indeed coming to an end, either there are no interesting questions left to answer or it is impossible to produce a set of scientific rules for answering any 'big question' that still piques our curiosity. It stretches the imagination to suppose that anyone could take either alternative seriously.

I am sure many scientists could reel off a lengthy list of major conundrums facing science today. How, for example, did life originate on Earth? Are human social behavioural patterns determined by our genes? How do humans acquire language? Is it possible to build a computing machine that will think like humans? Are there intelligent, extraterrestrial life forms in our Galaxy? Is there an objective reality independent of human observers? Even Horgan, who says that science is part of the "primordial human quest to understand the universe and our place in it", would probably agree that these questions are an integral part of that quest, and that the wellspring of deep and important questions

Strangely enough, I can't ever recall seeing a book or article suggesting that mathematicians are losing any sleep over the end of mathematics

is far from running dry. What is more, science is not much closer to offering a knockdown, airtight set of scientific rules for providing the answers than it was when these 'big questions' first arose. But that in no way implies that such a set of rules does not exist. An analogy here with similar 'big questions' in mathematics is helpful.

In 1931, Kurt Gödel squashed David Hilbert's cherished belief that any mathematical question could be definitively answered. Gödel's result demonstrated the existence of forever unanswerable questions about arithmetic. So here we have an area for which we can state unequivocally that there are questions that can never be answered by following the rules of mathematics. Yet, strangely enough, I can't ever recall seeing a book or article suggesting that mathematicians are losing any sleep over the end of mathematics. In fact, until recently, the undecidable propositions underwritten by Gödel's results were regarded mostly as curiosities by the mathematical community, although occasionally someone might start dreaming in print about one or another famous unsolved problem being one of them. The celebrated Fermat conjecture was once thought of in just these terms. What it took for the conjecture to be settled, however, was just a little more genius — and a lot more hard work — on the part of Andrew

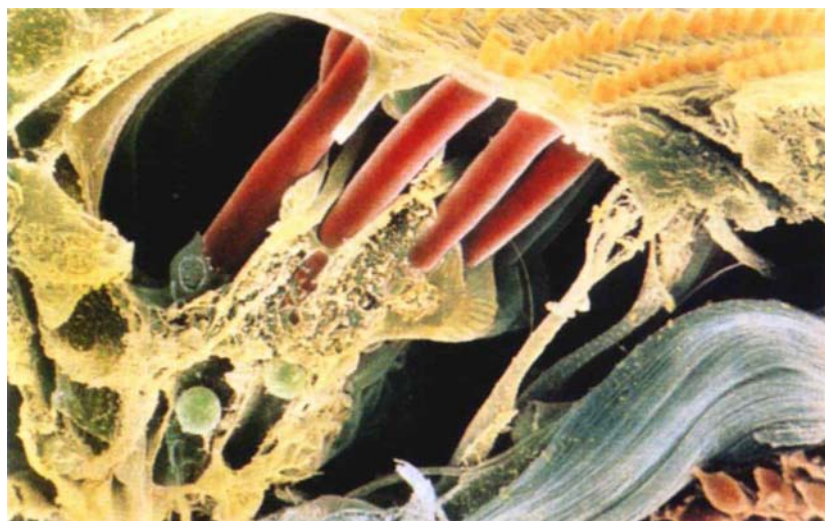
Wiles in wielding the traditional tools of mathematical argumentation.

Interestingly, while Horgan's chapter titles talk about the end of cosmology, evolutionary biology, physics and so on, there is no chapter on the end of mathematics. Perhaps this is because mathematics is often seen as mostly a process of inventing logically consistent worlds which Horgan claims is the way of the natural and biological sciences, instead of discovering them. I am sure many mathematicians would take issue with such an assessment of their art. Some even argue that the practice of mathematics is becoming more and more like that of empirical science, as we use the computer as a new type of laboratory to investigate the relationship between numbers, geometric forms and other abstract entities.

Changing the rules

The famed four-colour conjecture is a case in point. In 1976, Kenneth Appel and Wolfgang Haken showed that no more than four colours are needed to colour any planar map. In contrast to conventional mathematical proofs, which are at least in principle surveyable by the human mind, their result was based on the computational investigation of nearly 2,000 individual cases, involving many hundreds of hours of supercomputer calculations. Many mathematicians rejected this 'proof' because it did not play fair by the traditional rules of the games mathematicians play. Twenty years later we find that this computational exercise was merely the tip of an iceberg that might change the very rules of these games. The changes extend far into the mathematical realm, touching even on the foundation point of whether the traditional notions of what is meant by a 'proof' shouldn't be re-examined.

The same evolution of the rules of the game is just as likely to happen in science as in mathematics. All that is needed is a 'big question' requiring new concepts and new methods. For example, many systems constituting the warp and weft of everyday life — say a stock market or a traffic network — involve a collection of agents (traders or drivers) interacting on the basis of limited, local information. Moreover, these agents are intelligent and adaptive; their behaviour and interactions are determined by rules, just like those governing the behaviour of planets or molecules. But unlike these lifeless objects, adaptive agents are ready to change their rules in accordance with new information that comes their way, continually adjusting to their environment to prolong their own survival. So far, there is nothing remotely close to a formalism, or set of scientific rules, for even stating, let alone understanding, questions surrounding the weird and wondrous ways of such processes.



AURAL awe — section through the organ of Corti, in the human inner ear, showing four rows of hair cells (top right). Each cell contains around 100 individual hairs that translate sound into nerve impulses. SEM, $\times 1,500$. Taken from *Inside Information: Imaging the Human Body* by William A. Ewing. Thames and Hudson, £12.95 (pbk).

A few years ago, the Santa Fe Institute was formed to serve as a centre for the scientific investigation of complex adaptive systems. But the methods of choice for these studies are as different from the methods used in ordinary science as the use of the computer was to resolve the four-colour conjecture. Science, Santa-Fe-style, is based largely on the use of detailed computer simulations of real-world phenomena such as stock markets or the immune system. Such silicon surrogates provide a laboratory for carrying out controlled, repeatable experiments of the sort that are too expensive, too impractical, too time-consuming or simply too dangerous to do on the real-world system. And these are just the sorts of experiments the scientific method requires for the construction of a *scientific* theory of anything. Computers promise to change the frontiers of science fundamentally. For the first time in history, researchers will be able to observe phenomena such as the behaviour of a stock market under unusual economic circumstances or an ecosystem over several millennia of real time. And if history is any guide, such tools will generate a plethora of so far unstated 'big questions' that will serve as the basis for a bona fide science of complex systems in the decades to come.

One of Horgan's principle targets is exactly this claim. Part of his antipathy towards the development of a science of 'complexity' seems to be that such investigations do not emphasize the material and energetic structure of systems, the traditional foci of the scientific enterprise. Recalling Aristotle's theory of causation, one might say that the kinds of studies done at places such as the Santa Fe Institute are instead primarily concerned with

formal and final cause. So it is perhaps not surprising that Horgan (and others) find Santa-Fe-style science unconvincing, as it centres attention on issues of information and pattern that are basically orthogonal to the concerns of conventional science. But that is precisely why such investigations offer the potential of opening up whole new worlds of *real* science, rather than the so-called 'ironic' variant espoused by Horgan, which relies on vague personal opinions, subjective judgements on untestable hypotheses and semi-theological debate.

Unlike many of today's 'endologists' who hint darkly at the end of some field or other from their perspective as an active researcher in the area under scrutiny, journalistic members of the 'end-of-X' crowd have a predilection for invoking outside authority figures to buttress their claims. For some unaccountable reason, Nobel-prizewinning physicists seem especially popular in this regard. I, for one, am not sure that an eminent physicist, actively engaged in promoting his field, is the first person I would consult for a balanced, non-partisan view of the future of physics. Yet Horgan cites with benign approval Richard Feynman's remark that "[this] is the age in which we are discovering the fundamental laws of nature, and that day will never come again".

Let me appeal to the same shameless rhetorical trick in offering an antidote to Feynman's brand of misguided hubris. When told of the discovery of X-rays, Lord Kelvin, former president of the Royal Society, and one of the pre-eminent physicists of the late nineteenth century, solemnly intoned: "X-rays will prove to be a hoax." My friendly neighbourhood radiologist will no doubt ponder this point with great pleasure on his next trip to the bank. And on

his way from the bank to his summer home in the Swiss Alps, perhaps he will also reflect on another of Lord Kelvin's pronouncements: "I can state flatly that heavier-than-air flying machines are impossible." (Did Lord Kelvin ever see a bird?) All this brings to mind an observation by Arthur C. Clarke so pregnant with relevancy that it is now enshrined in the literature as Clarke's first law: "When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong."

One of the more appealing aspects of Horgan's book is its sheer entertainment value. The go-for-the-jugular writing style, coupled with the fact that just about everyone who's anyone in the scientific world has been subjected to one of Horgan's interviews, gives the reader an opportunity to see world-class scientists as "demigods on stilts", to use Einstein's pithy phrase describing the residents of Princeton, rather than monk-like keepers of the sacred flame. Horgan's jaundiced squint at these notables has no doubt led to the weeping and gnashing of teeth in some circles; nevertheless, it is not without some considerable charm—especially for those who have ever met any of Horgan's subjects. But just as remarks are not literature, accounts of personality quirks do not constitute philosophy. And while the book is amusing and the issues it raises are of great importance, its philosophical thrust and argumentation are just plain unconvincing.

There is one genuinely interesting point struggling to emerge from this whole debate. It is not whether science as we know it is coming to an end. Rather, it is whether the real world may not be just too complex for the human mind to comprehend fully. In other words, are there limits to what we can ever hope to know by using the tools and techniques of what we call 'science'? By 'limits' I mean specific 'big questions' that can be easily posed but never answered. If they do exist, I am sure we would all like to know about them. But unless, as Horgan would have it, they happen to encompass every 'big question' that we can conceive of asking about the world, we would still be as far away from the end of science as we were at its beginning.

Heavier-than-air flight is of course alive and well. Unfortunately, so too are lighter-than-air frothings about the end of science. So, despite Horgan's lively portrayal of modern science and scientists, what remains after the rhetorical flourishes all fade away is little more than a shapeless bit of intellectual fluff, pure cotton candy for the mind. □

John L. Casti is at the Santa Fe Institute, Santa Fe, New Mexico 87501, USA and the Technical University of Vienna, A-1040 Vienna, Austria. e-mail: casti@santafe.edu

Almost all about symmetry

Simon Altmann

Symmetries of Nature: A Handbook for Philosophy of Nature and Science. By Klaus Mainzer. Walter de Gruyter: 1996. Pp. 681. DM338.

ANAXIMANDER, the author of probably the first philosophy book, was the first to use a symmetry argument, according to Aristotle's *De Caelo*. This argument deduced the equilibrium of the Earth from the equi-distribution of the celestial bodies. It was the fountain-head of the great intellectual line that sprung from it, via Leibniz's "principle of sufficient reason", to Pierre Curie, who gave the first clear and complete enunciation of the "principle of symmetry".

All this and much more is in Klaus Mainzer's book. But readers cannot expect his overview to hit them in the eye. Mainzer's work has to be used more or less on the same principles as those articulated by Jorge Luis Borges' cartographers who produced a map of their country at a scale of 1 to 1. Everything is there, nothing is missed, but nothing is added either. So readers have to find their own route and even draw their own conclusions.

Take, for example, the third prong of Curie's principle, which states that the effects can be more symmetrical than the causes. (I am translating from the French, as the English version is seriously wrong, whether by mistake in the original or mis-translation from the German.) As Bertrand Russell said, the words 'cause' and 'effect' are not much used in science, but they happen to be part of the traditional vocabulary of symmetry theory. Scientists differ from many philosophers in that they never use the word 'cause' unless its absence implies the absence of the effect too (think about a force and its attendant acceleration). This is a condition that Curie's 'cause' does not satisfy, as follows from his rule as stated above.

One might have expected attention to this problem in the text, as this is claimed to be a book on the philosophy of science. But not a word on it will one find. Neither is there anything about the fact that the first two prongs of Curie's law are contrapositive and so logically equivalent. There is no discussion of the physical reasons that allow symmetry to be created where it did not exist, although this is a crucially important fact.

On the other hand, many topics are treated well. I particularly enjoyed the way in which Mainzer moves from the Platonic solids to their counterparts in four- and higher-dimensional spaces, and his approach to colour symmetry by

relating it to musical symmetries. And there is a great deal on Lie and Galois groups, and on symmetries in electromagnetism, relativity, classical and quantum mechanics, and elementary particles. I am not sure, however, why a Schrödinger cat has to be asphyxiated in a book on symmetry. But surely, if such activities are necessary, to dismiss the hapless animal as of "historical" value flies in the face of the evidence of the important, copious and entirely current literature on the subject.

As for relativity, I am not happy about the assertion that the Michelson-Morley experiment "confirms only the isotropy of the propagation of light". At least in the literature in English, isotropy is understood in this context to include the propagation of light along the same line in two opposite directions. The Michelson-Morley experiment would have saved a lot of theorists a great deal of sweat if it had been able to prove that these velocities are identical.

One serious problem with this book is that the German original is ten years old, and it shows. Mention is made of one of Jean Piaget's books on children's acquisition of the notions of movement and speed, without any reference to experiments by Jacques Mehler in Paris (already well known in the 1970s) that throw doubt on some of Piaget's results. Admittedly, Bas van Fraassen's admirable book *Laws and Symmetries* appeared too late for the German edition, but it is a great pity that the opportunity was not taken to use it when preparing the translation. (It might be unrealistic to expect such refinements in this edition, however, when no one seems to have thought it necessary to make sure that English books are cited in the original rather than in German translation.)

One of the most important philosophical problems raised by the principle of symmetry is that, badly used, it invites the delusion that symmetry allows for some sort of *a priori* knowledge. It is not sufficient in a book like this to say that this principle can be of "great heuristic use" subject to whatever caveats, if these caveats are not thoroughly spelt out. Poor Ørsted wasted eight years of work because of the wrong application of a symmetry principle, and he had a first-class mind.

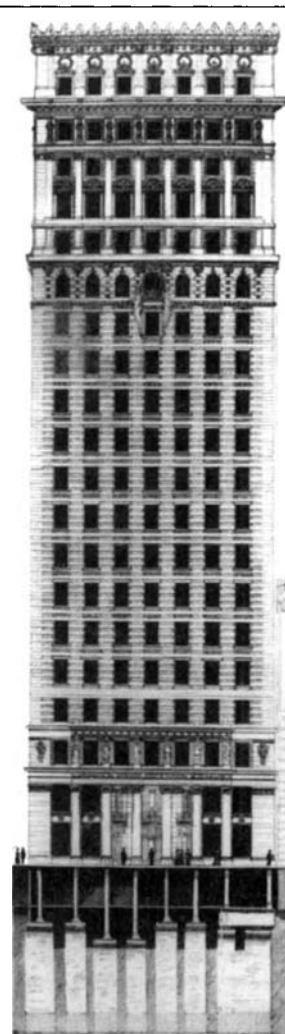
On the now established tradition in symmetry of the cultural grand tour, art is also included here but the choice of periods is idiosyncratic. Early art and even postmodern architecture are in, but the important problem of

right and left in the iconography of the Annunciation is not mentioned. And perhaps something about symmetry-breaking, as for instance by Antonio da Sangallo in the fenestration of the Palazzo Farnese in Rome, or by Francesco Borromini's glide plane in the balustrades of San Carlino, also in Rome, would have been appropriate. As for the postmodern period, surely the supreme example of symmetry-breaking is Hans Hollein's Haas Haus opposite St Stephen's Cathedral in Vienna.

But the most dramatic use of symmetry in architecture is the iconic one, which has given us the great divide between the style of Mies van der Rohe and the postmodern approach of his most famous pupil, Philip Johnson. This was forcefully displayed in 1979 by Johnson's broken pediment on the then AT&T building in Manhattan, which for many of us was the harbinger of the new order. Or is it disorder? Or is it a Kuhnian change of paradigm? It is a shame that the author does not discuss this, the most fascinating problem of present-day architecture. □

Simon Altmann is at Brasenose College, Oxford OX1 4AJ, UK.

On completion in 1894, the American Surety Building in New York was hailed as "a highly successful application of Greek detail to a modern office building" and the first of the city's pure "simplified" skyscrapers. A year later, a suit was brought against the American Surety Company charging it with obstructing the light and air of the building next door. In *Rise of the New York Skyscraper 1865-1913*, Sarah Bradford Landau, an architectural historian, and Carl W. Condit, a historian of technology, provide an impressively detailed history of New York's first skyscrapers while challenging conventional wisdom that it was in Chicago, not in New York, that the skyscraper was born. Yale University Press, \$50, £30.



In it for the eggs

T. R. Birkhead

Female Control: Sexual Selection by Cryptic Female Choice. By William G. Eberhard. Princeton University Press: 1996. Pp. 501. \$85, £60 (hbk); \$29.95, £24 (pbk).

FEMALES have always had a bad deal in evolutionary speculation. When Darwin first proposed the concept of sexual selection, he imagined two processes: competition between males, and female choice. It was obvious that males fought for access to females, but female choice was far from certain, and some people doubted whether females even have the mental ability to make such choices. It has taken more than a century of painstaking research to show that female choice is a subtle but important part of sexual selection.

Darwin and most subsequent students viewed sexual selection as the acquisition of partners through either fighting or female choice. But in 1970 biologists realized that there was more to sexual selection and that, even after copulation and insemination, competition between males could continue, through the process of sperm competition. Inside the passive female's reproductive tract, the sperm from different males grapple for paternity. Most behavioural ecologists were sexist, and it was more than a decade before the complementary idea of cryptic female choice emerged: that is, that females might also be able to influence the paternity of their offspring. But even this idea was too early for its time, and was generally thought of as barely credible. This was partly because it was far from clear how females could benefit from such choice: on theoretical grounds it seemed unlikely that they could improve the quality of their offspring.

A few years ago, a combination of events changed this perception. Field biologists noticed that it was females rather than males who initiated copulations with different males; theoreticians decided that it was possible for females to gain genetic benefits for their offspring by being fertilized by a particular male after all; and behavioural ecologists began to consider the mechanisms underlying reproductive processes instead of focusing exclusively on the adaptive significance of traits. Suddenly, the field was ripe for a female-centric view of reproductive competition.

William Eberhard's fascinating and scholarly review has finally provided just that. It is a landmark volume that marks a turning point in the study of sexual selection. By drawing our attention to the many different ways in which females may

potentially control paternity, the author opens up a whole new field of research.

But the way forward will not be easy, for it is full of traps. As this book reveals, the conflict between the sexes for the control of paternity has resulted in a bewildering array of subtle female anatomical and physiological adaptations, matched by equally remarkable male coevolutionary responses. Disentangling male-caused and female-caused variations in paternity and producing convincing demonstrations of cryptic female choice will be far from straightforward. Nevertheless, the potential rewards are considerable. What is more, acknowledging the existence of cryptic female choice forces us to reconsider the definition of sexual selection: males are not competing for females, they are competing for their ova. □

T. R. Birkhead is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

The three faces of Eve's therapist

Hugh Freeman

Alter Egos: Multiple Personalities. By David Cohen. Constable: 1996. Pp. 251. £18.95.

SINCE at least 1816 there have been reports of people apparently switching into different identities for periods from a few minutes to several weeks. Robert Louis Stevenson's *Dr Jekyll and Mr Hyde* illustrated the deep resonance of this possibility in the human mind. Later in the century in Paris, the investigation of hysteria by Jean-Martin Charcot and Pierre Janet, using hypnosis, appeared to show that dissociation of the ego could allow previously hidden aspects of an individual's personality to emerge. But this approach fell into disfavour when it was found that many of the remarkable changes seen in patients were actually induced by the therapists. Dissociation then remained little more than a clinical curiosity until the 1950s.

David Cohen's main concern is to describe the emergence since then of a new condition — multiple personality disorder — which has taken place almost entirely in North America. It is symbolized by the film *The Three Faces of Eve*, which was based on an early case and which has influenced patients and therapists in the United States for nearly 40 years. In fact, life seems to imitate art: Cohen reports that alternative personalities "often seem to model their performances on what they have seen on television". The main enthusiast for the

condition, Colin Ross, claims that more than six million Americans suffer from the disorder, but in the rest of the world only the Netherlands has recognized its existence to any significant extent.

What has turned the disorder into an important public issue is the belief of those who recognize and treat it that it is caused by abuse in childhood, usually sexual. That most subjects do not remember any such abuse is no problem to the therapists, as by intensive exploration, often under hypnosis, memories of it are 'recovered'. The next step may be for the patient — usually a young woman — to bring a court action against her parents, which may result in the father going to prison.

But there is, in fact, no scientific evidence whatever that multiple personality disorder — if such a condition exists — has any connection with abuse in childhood. 'False memory syndrome' is the name now given to the state in which 'memories' of abuse have been induced by the therapist, who then proceeds to 'treat' the alleged disorder. This practice can, of course, be very lucrative.

Cohen describes this imbroglio, but does not acknowledge the overwhelming harm it has done to individuals and families, some of whom will never recover. The issue was thoroughly debated last year in the correspondence columns of the *New York Review of Books*, although Cohen does not cite this reference.

But Cohen rightly points out that "the fact that memories have been repressed doesn't make them true". He has been unable to find any reliable follow-up studies of the treatment methods used by the therapists. "Ideally, talking through the trauma will pave the way for the creation of a new personality... It is increasingly clear, however, that relatively few cases have such happy endings."

Most likely, multiple personality disorder is a phenomenon that emerges when a suggestible subject comes into contact with a persistent therapist. Cohen suggests that high levels of stress often trigger episodes of dissociation, and that this may be one of the strange ways in which the unconscious speaks. Enthusiasts for the disorder are unlikely to be satisfied with that conclusion.

Alter Egos serves as an introductory account of multiple personality disorder for the nonspecialist reader, although it would have been strengthened by more scientific scepticism. □

Hugh Freeman, former editor of the *British Journal of Psychiatry*, is at 21 Montagu Square, London W1H 1RE, UK.

New Journals

Next week's issue will contain *Nature's* annual New Journals review supplement.

From plume head to continental lithosphere in the Arabian–Nubian shield

Mordechai Stein* & Steven L. Goldstein†‡

* Institute of Earth Sciences, The Hebrew University, Givat Ram, Jerusalem 91904, Israel

† Max Planck Institut für Chemie, Postfach 3060, D–55020 Mainz, Germany

The lithosphere of the Arabian–Nubian shield was mainly formed during an interval of about 150 million years near the end of the Proterozoic aeon. The events recorded in the rocks of the shield indicate that an oceanic plateau, formed by the head of an upwelling mantle plume, was later overprinted with continent-like characteristics by plate convergence and its associated magmatism. Similar sequences of events are seen in the geological record from Archaean to recent times, suggesting that the transformation from plume head to continental lithosphere has been an important component of continent generation throughout Earth history.

THE question of how continental lithosphere (the continental crust and its mantle root) is generated and stabilized is a fundamental problem in Earth science¹. It has been conventionally viewed as occurring through ‘arc’ magmatism at convergent plate margins, near the deep-sea trenches where old oceanic crust is subducted into the mantle². The arc crust itself reaches a convergent plate margin as a result of sea-floor spreading, usually within only tens of millions of years of its formation. This thick, silicic crust is sufficiently buoyant to resist subduction, and instead collides against a continental margin and with other arcs, resulting in the accretion of young continental lithosphere onto the edge of an older continental core. In the early 1980s it was suggested that shallow regions of the ocean floor, composed of submerged continental fragments, seamount chains and otherwise anomalously thick oceanic crust, will also resist subduction and may significantly contribute to the mass of accreted terranes^{3,4}. Reymer and Schubert^{5,6} later concluded that some continental segments, including the Arabian–Nubian and Canadian shields, were created at high rates which are difficult to explain by the arc accretion, and suggested that hotspot volcanism and underplating are important components of these provinces. Schubert and Sandwell⁷ quantitatively established the importance of oceanic plateaux as a fraction of the total oceanic and continental crust, and emphasized their importance for present-day continent accretion.

Many oceanic plateaux are generated during the earliest stage of ocean island volcanism. The ‘plume head’, formed through entrainment of large amounts of mantle during the initial stages of diapiric upwelling of a mantle plume, is many times larger than the rising mantle column associated with subsequent volcanics, and its arrival at the Earth’s surface is manifested by massive volcanic activity over a few million years^{8,9}. Strong evidence for a link between plume-head magmatism and the formation of a large continental mass was first shown for the early Proterozoic Birimian province of west Africa^{10,11}, and later suggested for the late Archaean Superior province of the Canadian shield^{12,13}. Nevertheless, the role of oceanic plateaux in continent formation is disputed. Multiple island arcs, if accreted together, can give a very high local crustal accretion rate, and recently the conventional arc accretion model has been used to explain the formation of large crustal segments, including the Superior province, the Arabian–Nubian shield (ANS), the North American cordillera and Eurasia^{14,15}.

The ANS is a large continental province affected by the late Proterozoic Pan African orogeny (loosely defined as occurring

1,200–550 Myr ago¹⁶). The conclusion^{5,6} that its growth rate was too high to be a result of arc accretion has been the subject of great debate^{14,17–21}. Based on stratigraphy, geochronology, secular changes in magma chemistry and neodymium, strontium and lead isotope ratios, we present evidence for anomalously rapid generation of both the ANS continental crust and its lithospheric mantle root, mainly over ~150 Myr, through formation of an oceanic plateau by a plume head, followed by arc magmatism along the plateau margins. On reaching convergent plate margins, oceanic plateaux are convenient nucleation points for arc magmatism because they resist subduction. Taking into account the link between plume-head magmatism and formation of the Birimian^{10,11} and Superior^{12,13} provinces, as well as the ANS, the evidence suggests that plume-head magmatism has been important in continent generation since the Archaean.

Formation of the Arabian–Nubian continent

Bentor²² divides the late Proterozoic ANS magmatic history into four phases. Thick sequences of oceanic tholeiitic basalts erupted from about 900 to 870 Myr during phase I, followed by island arc calc-alkaline magmatism until ~650 Myr during phase II. Pervasive metamorphism of the older rocks and the formation of presently unmetamorphosed granodiorite-to-granitic batholiths occurred between 640 and 600 Myr during phase III, followed by generation of alkaline granites and dolerites until ~540 Myr during the post-orogenic phase IV.

Strong isotopic evidence indicates that the core region of the ANS, from western Egypt to eastern Saudi Arabia, and from Israel and Jordan southwards to Sudan²¹, was generated as juvenile continental crust during the late Proterozoic. Stacey and Stoeser²³ delineated the boundaries of the juvenile Pan African regions of the ANS, in Saudi Arabia and Egypt, on the basis of lead isotopes in galenas, feldspars and whole rocks. Higher ²⁰⁸Pb/²⁰⁴Pb and ²⁰⁷Pb/²⁰⁴Pb ratios at similar ²⁰⁶Pb/²⁰⁴Pb ratios to the east and the west of this core region surrounding the Red Sea indicate a lead component from older continental crust. The Rb–Sr isochrons of island arc phase II magmas show low initial ⁸⁷Sr/⁸⁶Sr ratios of 0.703 ± 0.001 (Fig. 1), which preclude the significant involvement of an old, pre–900–Myr felsic continental precursor (refs 22, 24–27 and references therein). This is further supported by the mantle-like oxygen isotope ratios ($\delta^{18}\text{O} = +5.5$ to $+7$) of late Proterozoic rocks from Sinai²⁴. Further evidence for the absence of an old continental precursor is given by measurements of neodymium isotopes. Proterozoic magmas from the northern ANS show a small range of $\epsilon_{\text{Nd}}(t)$ (the deviation of ¹⁴³Nd/¹⁴⁴Nd ratios (in parts per 10⁴) from the bulk Earth value, when the rocks

† Present address: Lamont-Doherty Earth Observatory and Department of Earth and Environmental Sciences, Columbia University, Palisades, New York 10964, USA.

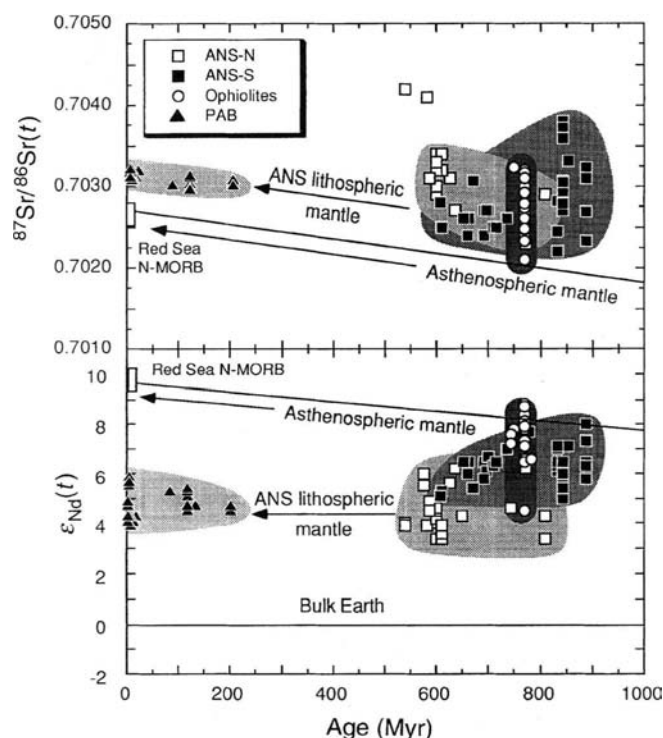


FIG. 1 Initial $^{87}\text{Sr}/^{86}\text{Sr}$ and ϵ_{Nd} values for the late Proterozoic (Pan African) ANS magmas, the Gabal Gerf ophiolite (representing late Proterozoic ocean ridge basalts (MORB)), Phanerozoic alkali basalts (PAB) from Israel and depleted Red Sea N-MORB. Gabal Gerf and Red Sea N-MORB reflect evolution of the MORB mantle. The low initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of late Proterozoic magmas rule out significant involvement of old continental crust. The ϵ_{Nd} of the late Proterozoic ANS magmas are lower (more enriched) than Proterozoic depleted MORB mantle. The Phanerozoic basalts have nearly the same initial ϵ_{Nd} and $^{87}\text{Sr}/^{86}\text{Sr}$ as ANS magmas, consistent with derivation from the same mantle source. This source is the lithospheric mantle, as it is associated with the ANS from the Proterozoic to the present day, and is distinct from the convecting upper mantle. ANS-N and ANS-S are the northern and southern parts of the shield, respectively. Data sources: this study and refs 24, 40, 53–58 and T. Reischmann (personal communication). Fields are drawn around data of the same type to aid visualization.

formed) of +3 to +6 (Table 1 and Fig. 1); some magmas from the southern ANS in Egypt and Sudan have a higher ϵ_{Nd} of +5 to +8 (Fig. 1). These values, which are lower than those for contemporaneous mid-ocean-ridge basalts (MORB) which represent 'depleted' mantle, but higher than those for the bulk Earth, indicate derivation from 'enriched' mantle or young continental sources. The occurrence of numerous ophiolites demonstrates tectonic associations with oceanic crust (ref. 21 and references therein).

Taken together, the initial Pb, Sr and Nd isotope ratios of late Proterozoic, ANS magmatic rocks, indicate formation during the late Proterozoic Pan African orogeny, in the absence of a significant felsic crustal precursor. Age relationships and magma compositions show that ocean basalt volcanism preceded convergent-margin magmatism.

Evolution of the ANS lithospheric mantle

The continental lithosphere consists of a silica-rich, low-density continental crust and a lithospheric mantle, isolated from large-scale mantle convective mixing. As a result, the isotope ratios of the lithospheric mantle evolve independently from the deeper, convecting asthenospheric mantle. Here we discuss the evidence that the ANS lithospheric mantle and continental crust formed from the same enriched mantle sources at the same time.

If rocks are formed at different times from a common source,

the younger rocks must have higher initial isotope ratios owing to radioactive decay. Low Rb/Sr ratios in mantle sources would result in small increases in the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Phanerozoic, rift-related alkali basalts, which have erupted since ~200 Myr ago in the northern ANS, have initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios similar to, or slightly higher than, the late Proterozoic Sinai basement magmas (Fig. 1). This is consistent with the derivation of both groups from the same mantle precursor with a low Rb/Sr ratio. Although this relationship alone does not establish a causal link, derivation from a common source is further supported by the nearly constant value of ϵ_{Nd} in Pan African and Phanerozoic magmas in the northern ANS (Fig. 1). We suggest that a common, enriched mantle source formed in the late Proterozoic, with a low Rb/Sr ratio and a nearly bulk Earth Sm/Nd ratio, the latter being higher than those for average continental crust but lower than those for MORB mantle sources. The positive initial ϵ_{Nd} values of this enriched mantle source show that before the late Proterozoic the precursor was, on average, depleted in incompatible elements (and therefore had Sm/Nd ratios higher than those for the bulk Earth). This sort of history is a common feature of enriched mantle sources, including those associated with many present-day ocean island basalts.

The Phanerozoic alkali basalt source can be isotopically distinguished from the convecting asthenospheric upper mantle

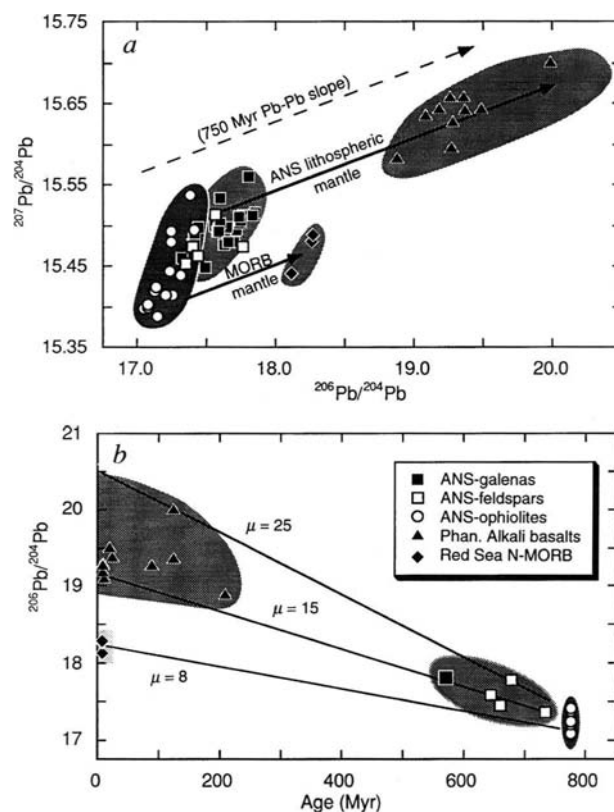


FIG. 2 Initial $^{206}\text{Pb}/^{204}\text{Pb}$ a, versus $^{207}\text{Pb}/^{204}\text{Pb}$, and b, against age, for late Proterozoic galenas and feldspars from the interior of the ANS, ANS ophiolites, Phanerozoic rift-related alkali basalts from Israel and Red Sea (depleted) N-MORB. The ophiolite samples and galenas display a range of $^{207}\text{Pb}/^{204}\text{Pb}$, interpreted as a mixing array between plume-head and depleted asthenospheric mantle components. Lines show Pb isotope evolution over 750 Myr. The evolution line from ophiolites to Red Sea N-MORB represents evolution of the convecting asthenospheric upper mantle. The line connecting galenas and feldspars with the Phanerozoic alkali basalts is consistent with evolution of the lithospheric mantle formed during the late Proterozoic along with the ANS continental crust. The divergence of $^{206}\text{Pb}/^{204}\text{Pb}$ ratios during the Phanerozoic reflects a much higher $^{238}\text{U}/^{204}\text{Pb}$ (≈ 20) in the lithospheric source of the Phanerozoic basalts compared with the asthenospheric upper mantle (≈ 8). In b, only samples of known age are plotted. Data sources: refs 23, 55, 57–59. Fields are drawn around data of the same type to aid visualization.

TABLE 1 Nd and Sr isotopes in late Proterozoic rocks from Sinai and southern Israel

Sample	Location*	Rock Type†	Age‡ (Myr)	Sm (p.p.m.)	Nd (p.p.m.)	$^{147}\text{Sm}/^{144}\text{Nd}$	$^{143}\text{Nd}/^{144}\text{Nd}$	ϵ_{Nd}	$\epsilon_{\text{Nd}}(T)$	Rb (p.p.m.)	Sr (p.p.m.)	$^{87}\text{Sr}/^{86}\text{Sr}\S$
ET-1	Eilat	Schist	800	3.602	19.65	0.1109	0.512366	-5.3	3.4	16	344	0.70587
ET-5A	Eilat	Schist	800	4.422	21.71	0.1232	0.512453	-3.6	3.9	46	335	0.70900
ET-5B	Eilat	Schist	800	4.488	21.90	0.1240	0.512493	-2.8	4.6	39	311	0.70836
ET-4	Eilat	Gr-Gneiss	744	6.281	32.35	0.1174	0.512503	-2.6	4.8	84	50	0.75579
RD-6	Roded	Qz Diorite	630	6.440	29.90	0.1303	0.512601	-0.7	4.6	29	998	0.70373
RD-10	Roded	Qz Diorite	630	6.825	32.29	0.1279	0.512590	-0.9	4.6			0.70422
RD-15x	Roded	Qz Diorite	630	5.133	20.54	0.1512	0.512640	0.0	3.7	10	993	0.70373
UM-2	Um-Maleq	Calc-alk. Gr	600	3.605	17.93	0.1216	0.512566	-1.4	4.3	118	185	0.71940
UM-9	Um-Maleq	Calc-alk. Gr	600	2.501	13.41	0.1128	0.512545	-1.8	4.6	205	153	0.73561
ET-20	Eilat	Calc-alk. Gr	600	1.809	8.902	0.1229	0.512534	-2.0	3.6	80	430	0.70781
ME-80	Ikna-Katherina	Alkaline-Gr	560	4.297	17.17	0.1514	0.512681	0.8	4.1	324	24.2	1.03600
M-2	Mandar	Alkaline-Gr	530	0.853	5.011	0.1029	0.512554	-1.6	4.7	236	16.8	0.00394
M-8	Mandar	Alkaline-Gr	530	0.927	4.692	0.1195	0.512581	-1.1	4.1	206	32.1	0.84170

Nd and Sr isotope ratios and Sm and Nd concentrations were measured on a Finnigan MAT 261 mass spectrometer. $^{143}\text{Nd}/^{144}\text{Nd}$ ratios are normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$, and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are normalized to $^{86}\text{Sr}/^{88}\text{Sr} = 0.1184$. During the period of study the values obtained for NBS 987 and La Jolla standards were $^{87}\text{Sr}/^{86}\text{Sr} = 0.710189$ ($2\sigma = 0.000022$, $n = 14$) and $^{143}\text{Nd}/^{144}\text{Nd} = 0.511850$ ($2\sigma = 0.000024$, $n = 24$), respectively. These are taken as the errors on the samples as they are larger than errors of individual measurements. Values are reported relative to $^{87}\text{Sr}/^{86}\text{Sr} = 0.710230$ for NBS 987 and $^{143}\text{Nd}/^{144}\text{Nd} = 0.511860$ for La Jolla. Chemical procedures are modified from ref. 50.

* Locations: Eilat and Roded massifs, southern Negev Desert, Israel; Um Maleq, Ikna Katherina and Mandar massifs, Sinai Peninsula, Egypt.

† Rock types: Schist, meta-graywacke with andesitic progenitor (phase II); Gr-Gneiss, calc-alkaline granitic gneiss (phase II); Qz-Diorite, Roded calc-alkaline quartz diorite (phase III); Calc-alk. Gr, calc-alkaline batholithic granites (phase III); Alkaline Gr, alkaline batholithic granites (phase IV).

‡ Rb-Sr isochron ages for granites and schist²⁴. Single zircon $^{207}\text{Pb}/^{206}\text{Pb}$ ages for granitic gneiss³⁹. The age of the Roded Quartz Diorite (RD-6) was determined using the single zircon Pb evaporation technique developed by Kober^{51,52}. The result was $^{207}\text{Pb}/^{206}\text{Pb} = 0.06081 \pm 0.0006$, corresponding to an age of 630 ± 3 Myr. The same age is assumed for samples RD-10 and RD-15x.

§ $^{87}\text{Sr}/^{86}\text{Sr}$ in UM-2, ME-80, M-2 and M-8 are from ref. 24.

beneath the ANS. The Arabian-Nubian continent has rifted to form the Red Sea, where depleted MORB, derived from the asthenospheric upper mantle, erupts. Depleted MORB from the Red Sea has a lower $^{87}\text{Sr}/^{86}\text{Sr}$ ratio and a higher ϵ_{Nd} value than the Phanerozoic alkali basalts (Fig. 1). The most depleted sources of ophiolites of Pan African age, with the highest initial ϵ_{Nd} and lowest $^{87}\text{Sr}/^{86}\text{Sr}$, are consistent with compositions expected for late Proterozoic MORB mantle. They can reasonably be linked with Red Sea depleted MORB, allowing for the radioactive decay of a common mantle source (Fig. 1). This evolution has a lower $^{87}\text{Sr}/^{86}\text{Sr}$ and a higher $\epsilon_{\text{Nd}}(t)$ than the other Pan African and Phanerozoic magmas, through time. The slopes of the $^{87}\text{Sr}/^{86}\text{Sr}$ and $\epsilon_{\text{Nd}}(t)$ curves give relative abundances of incompatible trace elements which imply that both sources had similar Rb/Sr ratios but the MORB source had higher Sm/Nd ratios. This means that the alkali basalt source is distinct from the asthenospheric mantle, both isotopically and chemically, but has been physically linked the ANS since the Proterozoic. We conclude that the source is lithospheric mantle, located above the asthenospheric mantle and isolated from convective mixing.

Lead isotope ratios are consistent with this interpretation, and give some age constraints on the lithospheric mantle. Pan African magmatic rocks, galenas and feldspars which are uncontaminated by older continental crust²³, indicate a mixing of components with low and high $^{207}\text{Pb}/^{204}\text{Pb}$ ratios (Fig. 2). Allowing for an increase in Pb isotope ratios because of U decay since 700–800 Myr, Phanerozoic alkali basalts can be linked to the high $^{207}\text{Pb}/^{204}\text{Pb}$ Pan African age component, and the Red Sea depleted MORB can be linked with the low $^{207}\text{Pb}/^{204}\text{Pb}$ component, which includes the ANS ophiolites (Fig. 2a). A common mantle source implies a $^{238}\text{U}/^{204}\text{Pb}$ ratio (μ) ≈ 8 (Fig. 2b). This is typical of the integrated μ value of the oceanic mantle over Earth history, which supports our suggestion that comparison of the late Proterozoic ophiolites and Red Sea depleted MORB represents the evolution of the convecting asthenospheric upper mantle. The lead isotopes in Phanerozoic alkali basalts are consistent with derivation from the same mantle source as Pan African galenas and feldspars after allowing for U decay since 700–800 Myr. The μ values of 15–25 are much higher than in the asthenospheric upper mantle (Fig. 2). The Pb isotope evolution, like that for Nd and Sr, is therefore consistent with our conclusion that the Phanerozoic alkali basalts are derived from the lithospheric mantle, which is located above

the asthenospheric mantle and is isolated from convective mixing.

The Sm–Nd isotopes in mantle xenoliths which are carried by Phanerozoic alkali basalts from Israel and Saudi Arabia^{28,29}, give additional age constraints on the lithospheric mantle. Incompatible element, depleted spinel lherzolite xenoliths, with $^{147}\text{Sm}/^{144}\text{Nd} \geq 0.19$, show a positive correlation of $^{147}\text{Sm}/^{144}\text{Nd}$ with $^{143}\text{Nd}/^{144}\text{Nd}$ (Fig. 3), consistent with an age of ~ 800 Myr and an initial ϵ_{Nd} of +5, and thus consistent with the timing of ANS crust generation and the $\epsilon_{\text{Nd}}(t)$ of ANS magmas. The major element chemistry is typical of peridotites which have undergone a major melt extraction, and those with higher Sm/Nd ratios are the least fertile²⁹. These data therefore suggest that a major melting event in the lithospheric mantle occurred the same time as the ANS continental crust was formed. Over one-third of the depleted xenoliths have $^{147}\text{Sm}/^{144}\text{Nd} = 0.18$ to 0.20 and $^{143}\text{Nd}/^{144}\text{Nd}$ similar to Phanerozoic alkali basalts. Enriched xenoliths, with $^{147}\text{Sm}/^{144}\text{Nd} \leq 0.18$, define a horizontal array which overlaps with the Phanerozoic alkali basalts. The similarity of the present-day isotope ratios of the basalts, the most fertile depleted xenoliths and the enriched xenoliths (Fig. 3), strongly suggests that the basalts are derived from the same lithospheric reservoir as the xenoliths (at greater depths). Moreover, the similarity of Nd isotope ratios in enriched xenoliths, despite a large range of Sm/Nd ratios, indicates that metasomatic events preceding incompatible element enrichment occurred recently, and are probably related to melt migration associated with Phanerozoic basaltic magmatism.

Although no single piece of evidence is definitive, the systematic isotopic evolution of ANS magma sources from the Proterozoic to the present day, the consistency of the coupled evolution of U–Pb isotope ratios with the formation age of the ANS in Proterozoic to recent samples, and the age–compositional relationships displayed by Sm–Nd isotopes in lithospheric mantle xenoliths, all strongly imply that the composition and age of the ANS lithospheric mantle can be characterized. Moreover, rift-related alkali basalts have erupted during the past 200 Myr, but show a small variation of $^{87}\text{Sr}/^{86}\text{Sr}$ and ϵ_{Nd} (Fig. 1) compared to the range of deep mantle magma sources, as shown by oceanic basalts³⁰. Such a small range of isotope ratios over such a long period of activity would be unlikely if the Phanerozoic alkali basalts were derived from a variety of deep mantle sources with different histories. The data indicate that the ANS lithospheric mantle and continental

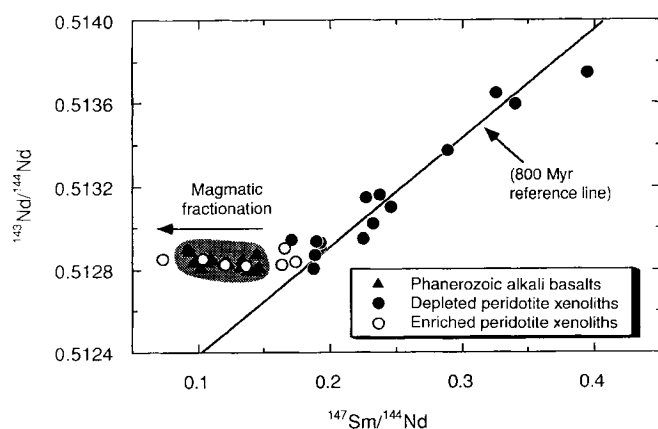


FIG. 3 $^{147}\text{Sm}/^{144}\text{Nd}$ against $^{143}\text{Nd}/^{144}\text{Nd}$ in lithospheric mantle-derived xenoliths from the ANS. The depleted peridotites ($^{147}\text{Sm}/^{144}\text{Nd} > 0.19$) lie on a positive correlation whose slope approximates to an 800 Myr reference isochron with an initial $\epsilon_{\text{Nd}} = +5$. These peridotites are residues of partial melting. The age is consistent with the timing of ANS calc-alkaline magmatism, and is interpreted as reflecting the time of melt extraction from the source region of the xenoliths. Many xenoliths cluster at $^{147}\text{Sm}/^{144}\text{Nd} = 0.18\text{--}0.20$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios similar to the Phanerozoic alkali basalts. These values appear to be a generally representative of the composition of the lithospheric mantle. Enriched xenoliths, with $^{147}\text{Sm}/^{144}\text{Nd} \leq 0.18$, define a horizontal array which overlaps with the Phanerozoic alkali basalts. The constancy of $^{143}\text{Nd}/^{144}\text{Nd}$ ratios, despite a large range of Sm/Nd ratios, indicates that metasomatic events leading to formation of the enriched xenoliths are recent, and probably related to melt migration associated with Phanerozoic basaltic magmatism. Data sources: refs 28, 29, 60.

crust both formed during the late Proterozoic from the same enriched mantle source, and the Phanerozoic basalts are derived primarily from this lithospheric mantle.

ANS crust formation rate

The evidence we have summarized shows that the ANS formed mainly during the late Proterozoic. Reymer and Schubert^{5,6} suggested that generation by arc magmatism would require an average arc addition rate of 310 km^3 per km arc length per Myr (assuming an area of $6 \times 10^6 \text{ km}^2$, a volume of $3 \times 10^8 \text{ km}^3$, a 300 Myr interval and an arc length of 2,500 km). This is an order of magnitude greater than the present-day average arc addition rate of $\sim 30 \text{ km}^3$ per km arc length per Myr. The suggested^{5,6} ANS accretion rate of $\sim 1 \text{ km}^3 \text{ yr}^{-1}$ equals the estimated^{5,6} total addition rate ($1.1 \text{ km}^3 \text{ yr}^{-1}$) for present-day arcs, and Reymer and Schubert concluded that other processes played a major role in ANS generation. This conclusion has been disputed on the grounds that the mass of juvenile, Pan African age continental crust¹⁷ was overestimated, or the length of convergent margins^{18,20} was underestimated, or both¹⁴. These later studies conclude that a smaller juvenile crustal mass and multiple active margins can result in more reasonable ANS accretion rate estimates explainable entirely by arc accretion, and may obviate any need for an additional accretion process.

Accretion rates are calculated based on the volume of continent generated and the length of the accretion interval. The above estimate^{5,6} assumes that the Pan African age basement extends from western Egypt to the Zagros mountains in Iran, and over a north-south length of 2,500 km. Dixon and Golombek¹⁷, assuming a Pan African crust formation age for only a narrow belt on either side of the Red Sea, give a volume estimate which is only one-fifth as large the estimate of Reymer and Schubert^{5,6}. However, subsequent studies agree that this estimate is too low^{18,20,21,31}, because Dixon and Golombek¹⁷ underestimate the area of the Pan African age continent and ignore the evidence for a mix of Pan African age and older continental crust on the margins of the ANS core region around the Red Sea. Stern²¹ recently estimated that the amount of

juvenile Pan African age crust was between 30 and 65% of the estimate of Reymer and Schubert^{5,6}. He considers his minimum estimate to be much too low because it includes only the area of outcropping juvenile, Pan African age crust and assumes that it contains $\sim 20\%$ reworked, older continental material. The maximum estimate adds the additional area affected by the Pan African orogeny and assumes that it contains a 30% juvenile component. Stern²¹ also considers the maximum estimate to be too conservative because in many places mapped as remobilized craton, as much as two-thirds of the material is juvenile. Moreover, there is evidence for much more widespread Pan African age continent generation near the ANS, in Oman³², as far west as Libya^{33,34} and as far south as Sudan and Somalia^{22,35}. It is thus likely that the amount of Pan African age continental crust associated with the ANS is similar to or exceeds the Reymer and Schubert estimate^{5,6}.

We suggest that previous ANS accretion rate estimates are biased toward low values because they assume a ~ 300 Myr interval, which is nearly equal to the total range of Pan African magma ages. The main continent generation phase occurred during oceanic and island arc phases (phases I and II), as the batholithic phase III involved significant remobilization of older continent, shown by the higher initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (> 0.7040) of many phase III magmas^{22,25}. Figure 4a shows the distribution of ages in magmas formed during phases I and II. If this reflects the true magmatic history, then the main crust formation activity occurred between about 900 and 750 Myr. The ANS age distribution is similar to that of the Birimian shield (Fig. 4) which has been recognized as a large continental province that accreted over a short period^{10,11}. Using Stern's²¹ volume estimates, this means that the accretion rate of the ANS alone was certainly greater than 60%, and possibly even greater than 130%, of the present-day global arc addition rate of $1.1 \text{ km}^3 \text{ yr}^{-1}$ (ref. 5). This strongly supports conclusion of Reymer and Schubert⁶, that the accretion rate of the ANS was anomalously rapid.

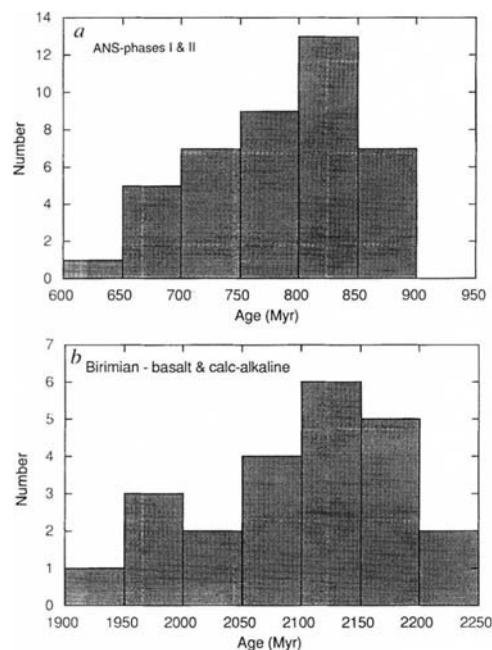


FIG. 4 Histogram of ages of a, ANS (phases I and II) and b, Birimian magmas. Both diagrams show similar clustering of ages during the first 150 Myr of magmatic activity, and decay of magma formation rates with time, culminating in the termination of subduction process (~ 650 Myr in the ANS). We relate phase I tholeiites to formation of an oceanic plateau and phase II to convergent margin magmatism. The Birimian diagram includes ages from the Guyana shield. Data sources: a, refs 24, 35, 36, 53, 54, 56, 61–67 and T. Reischmann (personal communication); b, refs 10, 11, 68.

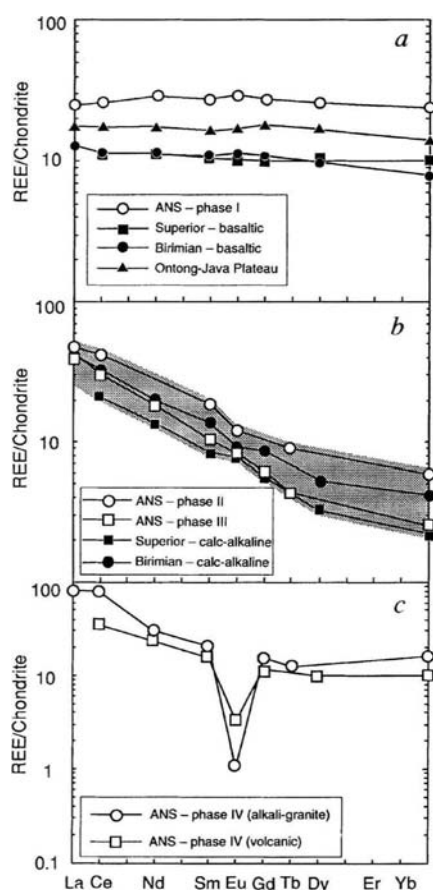


FIG. 5 Rare-earth element (REE) patterns from different magmatic phases of the ANS. *a*, Comparison of tholeiitic basalts from the oceanic phase I of ANS magmatism with tholeiites from the Ontong Java Plateau, the 2.1 Gyr Birimian orogen, and the 2.7 Gyr Superior province. All have similar, flat REE patterns. The Ontong Java Plateau is the largest on Earth, and the ANS, the Birimian and Superior province tholeiites are thought to represent oceanic plateaux generated by plume heads¹⁰⁻¹³. *b*, Calc-alkaline phase II mafic volcanics and batholithic phase III plutonics of the ANS, calc-alkaline magmas from the Birimian and Superior provinces. All are similar and light-REE and enriched. The ANS phases II and III mafic magmas lack significant Eu anomalies. *c*, Phase IV alkaline granites and alkaline volcanics have strong negative Eu anomalies (and low Sr concentrations) indicating fractionation of plagioclase. They appear to be fractionation products of mafic magmas formed in the lithospheric mantle. Data sources: refs 10, 37, 37, 42, 43, 49, 69, 70.

From plume head to continental crust

The high crust formation rate in the ANS indicates that processes other than arc magmatism played a major role in its generation. Data on 'oceanic' phase I lavas (mainly tholeiitic basalts) are consistent with its initial formation as an oceanic plateau. The Baish group in Saudi Arabia is ~16 km thick, which is too thick to represent normal oceanic crust²². Although some tectonic thickening cannot be precluded, it is not mentioned in current descriptions²², and compared to normal oceanic crust an unreasonably high thickening factor of 5 to 8 would be required. An oceanic setting is also indicated by phase I sediments which are immature volcanogenic clastics and breccias without signs of continental provenance (ref. 22 and references therein). Rare-earth element (REE) patterns of phase I basalts are nearly flat, and this is similar to the patterns of lavas from the Ontong Java Plateau (Fig. 5), an oceanic plateau in the western Pacific which formed by a plume head. The phase I basalt REE patterns are also similar to those of the oldest basalts from the ~2-Gyr-old Birimian shield, and those of basalts and komatiites of the ~2.7-Gyr-old Superior province.

Island arc phase II magmatism began at ~850 Myr, some 30–

50 Myr after phase I^{22,36}, which probably coincides with arrival of the oceanic plateau at a convergent plate margin. This phase was its most concentrated over the next ~100 Myr (Fig. 4). Magma compositions include calc-alkaline andesites, dacites and diorite-tonalite-trondhjemites (ref. 22 and references therein). The chemical transition from phase I to II, that is, of early tholeiitic basalts or komatiites superceded by calc-alkaline andesites and dacites and plutonic equivalents, is also seen in the Birimian^{10,11} and Superior¹² provinces. Phase II calc-alkaline magmas show enriched light-REE patterns and fractionated heavy-REE patterns, with weak negative Eu anomalies (Fig. 5). Stern³⁷ suggested that the Young Metavolcanics of Egypt are a fractional fusion product of hydrous garnet lherzolite. His model requires a flat REE pattern for the parent, which is consistent with the bulk Earth Sm/Nd value which we infer for the source of the Phanerozoic alkali basalts (Fig. 1), and is also consistent with the flat REE patterns of phase I tholeiites (Fig. 5). The large volume of calc-alkaline magmas emplaced during phase II, might be related to a thermal anomaly in the uppermost mantle generated by the plume head.

Phases III and IV represent the culmination of Pan African continent formation. The batholithic phase III is marked by lithospheric thickening, major metamorphism at ~650 Myr and the generation of voluminous calc-alkaline granitoids and minor gabbros and diorites, until 600 Myr. Even mafic examples display steep REE patterns resembling tonalities and dacites of other major Archaean and Proterozoic orogens (Fig. 5). Phase III represents the agglomeration of ANS terranes and collision with older cratonic blocks to the east and west^{21,38,39}. This was followed by the formation of alkaline granites and dolerites between 600 and 540 Myr during phase IV. Alkaline granites have strong negative Eu anomalies (Fig. 5) and low Sr concentrations (refs 24, 40, 41 and Table 1), indicating extensive plagioclase fractionation within the crust. Magmas appear to be crustal fractionation products of mafic lithospheric mantle melts⁴¹⁻⁴³. The alkaline magmatism is associated with normal faulting and volcanoclastic sedimentation. The transition from compressional to extensional tectonics was associated with relaxation of the lithosphere and probably the delamination of its lower part⁴⁴.

Plumes and continental growth

We have delineated a series of events which produced the ANS continental crust and lithospheric mantle. The main phase of crust formation occurred at an anomalously high rate between about 875 and 725 Myr (Fig. 4a), through oceanic and then calc-alkaline magmatism. The chemistry and stratigraphic relationships of magmas of Pan African age and their relationship to the Phanerozoic magmatism, indicate initial formation of a large lithospheric mass by a plume head which erupted on oceanic crust and generated an oceanic plateau. Upon reaching a convergent plate margin, the thick oceanic plateau lithosphere resisted subduction, and plate convergence occurred on its own margins, generating calc-alkaline magmas (Fig. 4) which may be partly a result of the thermal anomaly in the upper mantle caused by the plume head. Collision of the continental blocks was accompanied by regional metamorphism and the formation of batholiths with high initial ⁸⁷Sr/⁸⁶Sr ratios. The transition after 600 Myr from calc-alkaline magmatism to alkaline magmatism, the latter being associated with lithospheric mantle melting, reflects the end of the plate convergence following agglomeration.

This history of the ANS can be compared with the early Proterozoic Birimian shield in west Africa, an example of rapid continent formation initiated by the formation of an oceanic plateau^{10,11}. As in the ANS, the earliest magmas are thick sequences of tholeiitic basalts, but surface outcrops throughout the shield are mainly calc-alkaline. The main magmatic phases occurred over a short period of ~125 Myr and the age pattern is remarkably similar to that of the ANS (Fig. 4). Birimian magmas have $\epsilon_{\text{Nd}}(T) \approx +3$ indicating enriched mantle sources as in the ANS, and initial ⁸⁷Sr/⁸⁶Sr are less than 0.703 (refs 10, 11). These similarities, in terms of the geological sequence of events, the

interval of crust generation and the compositions of the sources, suggest that they underwent the same crust formation processes.

Using the Phanerozoic accretion of the North American cordillera as an analogue, Samson and Patchett¹⁴ suggested that the ANS formed through the accretion of multiple island arc terranes as a result of plate convergence and transcurrent motion. In order to support an arc accretion model, they suggest that the accretion rate is not unusually rapid, assume a long (300 Myr) formation interval and favour the low ANS volume estimate of Dixon and Golombek¹⁷. Initial Nd, Sr and Pb isotope ratios of ANS magmas (Figs 1 and 2), between MORB and old continental crust, can be compatible with their model as well as with ours. In many island arcs, comparable isotopic compositions reflect mixtures of depleted mantle and subducted sediments⁴⁵. However, present-day island arcs show a large variation of Nd and Sr isotope ratios⁴⁶. The range in the ANS, on the other hand, is rather restricted and comparable to those of other major orogens which appear to be associated with oceanic plateau volcanism^{10,11,47}, and is similar to the range in the Ontong Java and Manihiki oceanic plateaux^{48,49}.

Although a strict interpretation of the arc-accretion model is at odds with our interpretation, in many respects the differences are a matter of emphasis. Both models can result in high local crust accretion rates, and both acknowledge the importance of arc magmatism in the formation of continental crust, which in the history of the ANS, corresponds to phase II. The ANS contains transverse sutures which may indicate multiple terranes. Samson and Patchett¹⁴ acknowledge, but do not emphasise the role of oceanic basaltic magmatism in both provinces. For example, Wrangellia, one of the largest terranes in the North American cordillera, contains tholeiitic basalts greater than 6 km thick, which probably formed as an oceanic plateau. The Cache Creek terrane

has also been interpreted as an accreted oceanic plateau (refs 3, 4, 14 and references therein). Thus available evidence indicates that plume heads are important in the formation of the Cordillera.

Moreover, oceanic plateaux may well be an important component of present-day arc magmatism. Many present-day oceanic convergent plate margins are either directly associated with or located close to anomalously thick oceanic crust. These include the central Aleutians (associated with the Umnak Plateau and the Bowers and Shirshov ridges), Tonga, New Hebrides and Fiji (which have plateau or plume associations) and the Lesser Antilles (located near the site of the Caribbean plume head). Many of the rest, including the western Aleutians, the Kuriles and Kermadec, lie between long stretches of convergence associated with continent or oceanic plateaux. In fact, arc segments without some clear association with large igneous provinces, like the Marianas and Bonin arcs, are the exception rather than the rule.

The sequence of events in the ANS, involving formation of anomalously thick oceanic lithosphere modified by subduction at its margins, explains both the formation of major continental segments during discrete, short episodes and their calc-alkaline affinities. Stein and Hofmann⁴⁷ proposed that the Pan African orogeny and other major crustal formation episodes are related to mantle overturn/major orogeny (MOMO) events, which replenish the upper mantle in trace elements and are characterized by large scale plume activity. The magmatic history of the ANS may represent a complete MOMO cycle. Generation of large amounts of lithosphere over short periods, in association with plume-head magmatism, appears to be an important mechanism of continental growth in other regions of the Earth^{10-13,47}. The process appears to be the primary method of continental crust formation and cratonization throughout the history of the Earth, and is particularly important in cases of rapid continental growth. □

Received 31 May; accepted 26 July 1996.

- Rudnick, R. *Nature* **378**, 571–578 (1995).
- Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* 1–312 (Blackwell, Oxford, 1985).
- Ben-Avraham, Z., Nur, A., Jones, D. & Cox, A. *Science* **213**, 47–54 (1981).
- Nur, A. & Ben-Avraham, Z. *J. Geophys. Res.* **87**, 3644–3661 (1982).
- Reymer, A. & Schubert, G. *Tect.* **3**, 63–77 (1984).
- Reymer, A. & Schubert, G. *Geol.* **14**, 299–302 (1986).
- Schubert, G. & Sandwell, D. *Earth. Planet. Sci. Lett.* **92**, 234–246 (1989).
- Campbell, I. H. & Griffiths, R. W. *Earth Planet. Sci. Lett.* **99**, 79–93 (1990).
- Griffiths, R. W. & Campbell, I. H. *Earth Planet. Sci. Lett.* **99**, 66–78 (1990).
- Abouchami, W., Boher, M., Michard, A. & Albarède, F. *J. Geophys. Res.* **95**, 17605–17629 (1990).
- Boher, M., Abouchami, W., Michard, A., Albarède, F. & Arndt, N. T. *J. Geophys. Res.* **97**, 345–369 (1992).
- Desrochers, J.-P., Hubert, C., Ludden, J. N. & Pilote, P. *Geol.* **21**, 451–454 (1993).
- Kimura, G., Ludden, J. N., Desrochers, J.-P. & Hori, R. *Lithos* **30**, 337–355 (1993).
- Samson, S. D. & Patchett, P. *J. Aust. J. Earth Sci.* **38**, 595–611 (1991).
- Sengör, A. M. C., Natal'in, B. A. & Burtman, V. S. *Nature* **364**, 299–307 (1993).
- Jackson, N. J. & Ramsay, C. R. *Geol.* **8**, 210–211 (1980).
- Dixon, T. H. & Golombek, M. P. *Geol.* **16**, 991–994 (1988).
- Harris, N. B. W., Gass, I. G. & Hawkesworth, C. J. *Phil. Trans. R. Soc. Lond. A* **331**, 533–548 (1990).
- Sultan, M. *et al. Geol.* **18**, 761–764 (1990).
- Pallister, J. S., Cole, J. C., Stoesser, D. B. & Quick, J. E. *Geol.* **18**, 35–39 (1990).
- Stern, R. J. *Annu. Rev. Earth Planet. Sci.* **22**, 319–351 (1994).
- Bentor, Y. *Precamb. Res.* **28**, 1–74 (1985).
- Stacey, J. S. & Stoesser, D. B. *Contrib. Mineral. Petrol.* **84**, 91–105 (1983).
- Bielski, M. thesis, Hebrew Univ. Jerusalem (1982).
- Duyverman, H. J. & Harris, N. B. W. *Bull. Geol. Soc. Am.* **93**, 174–178 (1982).
- McGuire, A. V. & Stern, R. J. *Contrib. Mineral. Petrol.* **114**, 395–408 (1993).
- Stern, R. J. & Hedge, C. E. *Am. J. Sci.* **285**, 97–127 (1985).
- Henjes-Kunst, F., Altherr, R. & Baumann, A. *Contrib. Mineral. Petrol.* **105**, 460–472 (1990).
- Stein, M., Garfunkel, Z. & Jagoutz, E. *Geochim. Cosmochim. Acta* **57**, 1325–1337 (1993).
- Zindler, A. & Hart, S. *Annu. Rev. Earth Planet. Sci.* **14**, 493–571 (1986).
- Sultan, M., Bickford, M. E., El Kailoubi, B. & Arvidson, R. E. *Bull. Geol. Soc. Am.* **104**, 456–470 (1992).
- Gass, I. G., Shackleton, R. M. & Smewing, J. D. in *The Geology and Tectonics of the Oman Region* (eds Robertson, A., Searle, M. & Ries, A.) 585–599 (Geological Society of London, Edinburgh, 1990).
- Fullagar, P. D. in *Geology of Libya* (eds M. J. Salem & Busirewili, M. T.) 1051–1058 (Academic Press, New York, 1980).
- Bertrand, J. M. L. & Cabry, R. *Geol. Rundsch.* **67**, 357–388 (1978).
- Reischmann, T. & Kröner, A. *Geol. Rundsch.* **83**, 547–563 (1994).
- Reischmann, T., Bachtadse, V., Kröner, A. & Layer, P. *Earth. Planet. Sci. Lett.* **114**, 1–15 (1992).
- Stern, R. J. *Precamb. Res.* **16**, 195–230 (1981).
- Kröner, A. *Precamb. Res.* **27**, 277–300 (1985).
- Kröner, A., Eyal, M. & Eyal, Y. *Geol.* **18**, 545–548 (1990).

- Beyth, M., Stern, R. J., Altherr, R. & Kröner, A. *Lithos* **31**, 103–124 (1993).
- Kessel, R. thesis, Hebrew Univ. Jerusalem (1995).
- Stern, R. J. & Gottfried, D. *Contrib. Mineral. Petrol.* **92**, 492–501 (1986).
- Beyth, M., Stern, R., Altherr, R., Peltz, M. & Heiman, A. *Geol. Surv. Israel Curr. Rep.* **9**, 24–26 (1994).
- Black, R. & Liegeois, J. P. *J. Geol. Soc. (Lond.)* **150**, 89–98 (1993).
- Miller, D. M., Goldstein, S. L. & Langmuir, C. H. *Nature* **368**, 514–520 (1994).
- Hawkesworth, C. J., Gallagher, K., Hergt, J. M. & McDermott, F. *Annu. Rev. Earth Planet. Sci.* **21**, 175–202 (1993).
- Stein, M. & Hofmann, A. W. *Nature* **372**, 63–68 (1994).
- Mahoney, J. J. & Spencer, K. J. *Earth Planet. Sci. Lett.* **104**, 196–210 (1991).
- Mahoney, J. J., Storey, M., Duncan, R. A., Spencer, K. J. & Pringle, M. in *The Mesozoic Pacific: Geology, Tectonics and Volcanism* (eds Pringle, M. S., Sager, W. W., Sliter, W. V. & Stein, S.) 233–261 (Am. Geophys. Union, Washington, 1993).
- White, W. M. & Patchett, P. J. *Earth Planet. Sci. Lett.* **67**, 167–185 (1984).
- Kober, B. *Contrib. Mineral. Petrol.* **93**, 482–490 (1986).
- Kober, B. *Contrib. Mineral. Petrol.* **96**, 63–71 (1987).
- Duyverman, H. J., Harris, N. B. W. & Hawkesworth, C. J. *Earth Planet. Sci. Lett.* **59**, 315–326 (1982).
- Claesson, S., Pallister, J. S. & Tatsumoto, M. *Contrib. Mineral. Petrol.* **85**, 244–252 (1984).
- Stein, M. & Hofmann, A. W. *Earth Planet. Sci. Lett.* **114**, 193–209 (1992).
- Stern, R. J. & Kröner, A. *J. Geol.* **101**, 555–574 (1993).
- Volker, F., McCulloch, M. T. & Altherr, R. *Geophys. Res. Lett.* **20**, 927–930 (1993).
- Zimmer, M., Kröner, A., Jochum, K. P., Reischmann, T. & Todt, W. *Chem. Geol.* **123**, 29–51 (1995).
- Stacey, J. S., Doe, B. R., Roberts, R. J., Delevaux, M. H. & Gramlich, J. W. *Contrib. Mineral. Petrol.* **74**, 175–188 (1980).
- Blusztajn, J., Hart, S. R., Shimizu, N. & McGuire, A. V. *Chem. Geol.* **123**, 53–65 (1995).
- Fleck, R. J., Greenwood, W. R., Hadley, D. G., Anderson, R. E. & Schmidt, D. L. *US Geol. Surv. Prof. Pap.* **1131**, 1–38 (1980).
- Kröner, A., Byerly, G. R. & Lowe, D. R. *Earth Planet. Sci. Lett.* **103**, 41–54 (1991).
- Kröner, A., Pallister, J. S. & Fleck, R. J. *Geol.* **20**, 803–806 (1992).
- Halpern, M. & Pfister, N. J. *Geol.* **89**, 639–648 (1981).
- Darbyshire, D. P. F., Jackson, N. J., Ramsey, C. R. & Roobol, N. J. *J. Geol. Soc. (Lond.)* **140**, 203–213 (1983).
- Dixon, T. H. *Precamb. Res.* **14**, 119–133 (1981).
- Pallister, J. S., Stacey, J. S., Fischer, L. B. & Premo, W. R. *Precamb. Res.* **38**, 1–54 (1988).
- Gruau, G., Martin, H., Leveque, B. & Capdevila, R. *Precamb. Res.* **30**, 63–80 (1985).
- Arth, J. G. & Hanson, G. N. *Geochim. Cosmochim. Acta* **39**, 325–362 (1975).
- Reischmann, T., Kröner, A. & Basahel, A. *Bull. Fac. Earth Sci. King Abdul Aziz Univ.* **6**, 365–378 (1983).

ACKNOWLEDGEMENTS. This work was inspired by Y. K. Bentor, whose years of pioneering work in Sinai and southern Israel played a major role in revealing the stratigraphy of the ANS. We thank F. Albarède, R. Bohannon, and J. Patchett for reviews which led to improvements in the manuscript; A. Kröner and T. Reischmann for discussions and help with zircon evaporation; D. Avigad for help in the field and M. Bielski for providing samples. M.S. was partly supported at MPI–Mainz by a Minerva Fellowship.

CORRESPONDENCE should be addressed to M.S. (e-mail: motis@vms.huji.ac.il)

Transmission dynamics and epidemiology of BSE in British cattle

R. M. Anderson*, C. A. Donnelly*, N. M. Ferguson*, M. E. J. Woolhouse*, C. J. Watt*, H. J. Udy*, S. MaWhinney†, S. P. Dunstan*, T. R. E. Southwood‡, J. W. Wilesmith§, J. B. M. Ryan§, L. J. Hoinville§, J. E. Hillerton||, A. R. Austin§ & G. A. H. Wells§

* Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

† University of Colorado Health Sciences Center, Department of Preventative Medicine and Biometrics, Campus Box B119, 4200 East Ninth Avenue, Denver, Colorado 80262, USA

‡ Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

§ Central Veterinary Laboratory, New Haw, Addlestone, Surrey KT15 3NB, UK

|| Institute of Animal Health, Compton, Berkshire NG20 7NN, UK

A comprehensive analysis of the bovine spongiform encephalopathy (BSE) epidemic in cattle in Great Britain assesses past, present and future patterns in the incidence of infection and disease, and allows a critical appraisal of different culling policies for eradication of the disease.

In March 1996 the Spongiform Encephalopathy Advisory Committee (SEAC) of the Department of Health (DoH) and Ministry of Agriculture, Fisheries and Food (MAFF) in the United Kingdom reported to government that the most likely explanation at present for the occurrence of 10 cases (now 12 confirmed cases)^{1,2} of an apparently new variant of the rare but lethal Creutzfeldt-Jakob disease (CJD) in humans was exposure to the aetiological agent of BSE (believed to be an abnormal form of the prion protein (PrP)³) before regulations were introduced to prevent any part of cattle with clinical signs (1988) and specified offal from all cattle entering the human food chain (1989)^{4,5}. This announcement triggered a crisis in Europe, both in consumer confidence in beef and beef products, and in relations between the member states of the European Union.

An analysis is presented of the key epidemiological processes that have shaped the pattern of the BSE epidemic in Great Britain, and of the demography of the cattle population, using new data on the incubation period of the disease and the rate of maternal transmission from infected dam to a new-born calf⁶. Mathematical models are used to derive estimates of the age-stratified incidence of infection in cattle over time and the likely impact of various culling policies⁷ on the incidence of BSE in the coming years.

Origins and time course of the epidemic

Current evidence suggests that the disease originated from supplementary feed containing meat and bone meal (MBM) contaminated by a scrapie-like agent derived from sheep or cattle (the oral route of infection for scrapie-like agents has been demonstrated experimentally^{8,9}). Infectivity is thought to be facilitated by changes in the early 1980s in the method of rendering of offal into meat and bone meal¹⁰. Legislation was introduced in June 1988 to make BSE a notifiable disease. Shortly afterwards a statutory ban on the feeding of ruminant-derived protein to ruminants and regulations for the compulsory destruction of all animals exhibiting signs¹¹ were introduced. By June 1996, 161,412 confirmed cases of BSE had been reported in Great Britain. The pattern of the epidemic remains consistent with the hypothesis that the vast majority of cases arose by infection with contaminated feed (feed-borne hypothesis)¹²⁻¹⁶. It remains possible, however, that other routes of transmission may occur infrequently, in particular maternal transmission from dam to calf⁶.

In contrast to studies of transmissible spongiform encephalo-

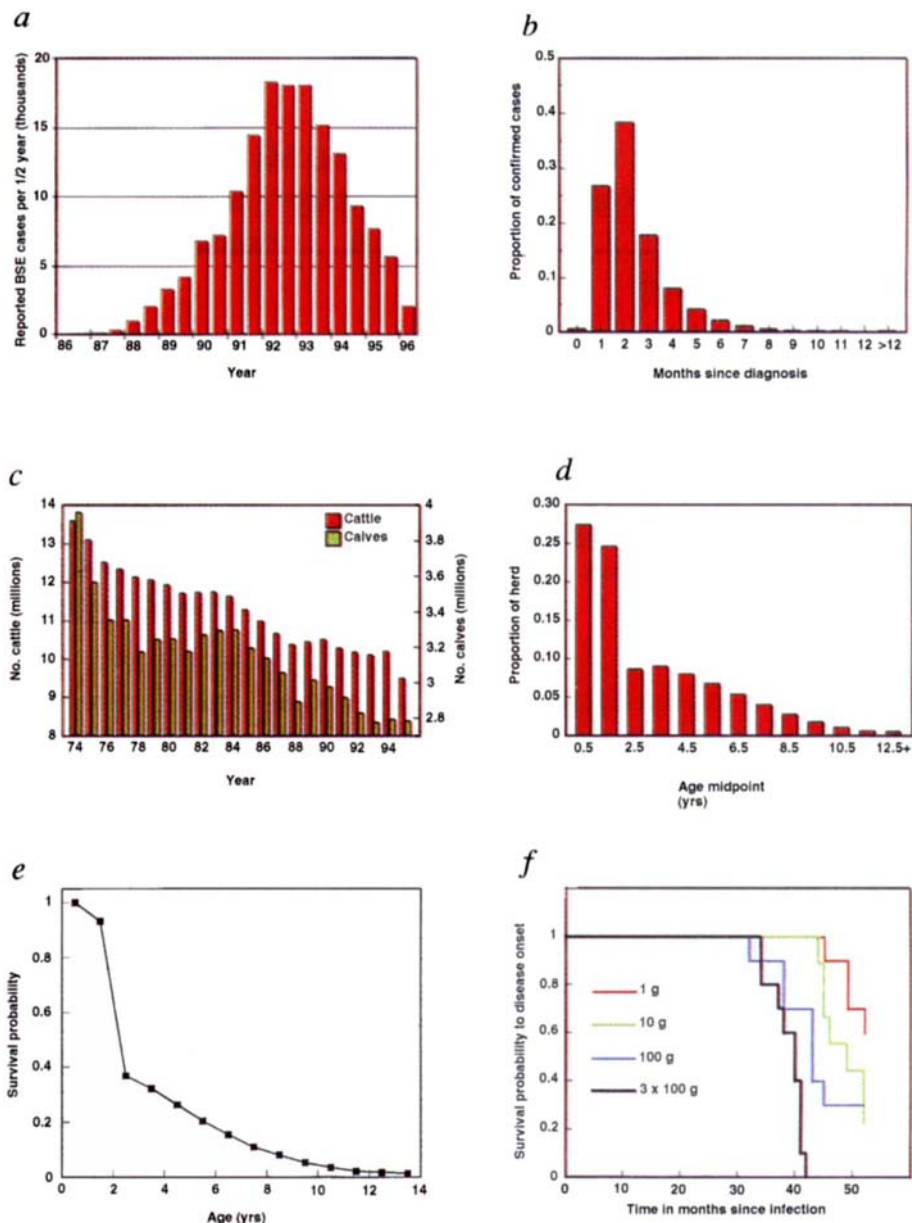
pathies (TSEs) in rodent models¹⁷, sheep and goats¹⁸⁻²⁰, to date there is no evidence that cattle genotype influences the susceptibility or average duration of the incubation period of BSE²¹⁻²³, although this remains a possibility. There is much geographical heterogeneity in the incidence of BSE per head of cattle in Great Britain (Fig. 1g), probably because of variation in the past use of meat and bone meal in supplementary feedstuffs, geographical variation in the proportion of meat and bone meal produced by rendering plants, and variation in the use of hydrocarbon-solvent extraction procedures (the use of which is thought to reduce the likelihood of agent survival)¹⁰.

The observed epidemic pattern is influenced by under-reporting before the disease became notifiable in June 1988 (a degree of under-reporting may have continued for one or two years after this) and the short average reporting delay (approximately 3 months) which influences the most recent records (Fig. 1b). Our analyses exclude recent reports influenced by the delay. The rapid growth of the epidemic from 1986 to 1991 is consistent with recycling of contaminated material in the ruminant feed chain fuelling transmission before July 1988. The lag after this period before the incidence of disease decays is consistent with the hypothesis of a long average incubation period (4-6 years)^{24,25} (Fig. 1a). Examination of the age distribution of the incidence of disease in cohorts of animals of defined birth years, under the assumption that most animals acquire infection soon after birth, suggests a mean incubation period of approximately 5 years, with marked variation around the average but little variation between cohorts (Fig. 2a). An inspection of the cohort data on proportional incidence by age (Fig. 2b) reveals a trend for the average age at clinical onset to decrease within the cohorts born in 1977 to those born more recently. This pattern provides evidence that infection of older animals (>5 years of age) is possible through contaminated feed²⁶. It is also consistent with the hypothesis of recycling fuelling the pace of the epidemic, where a raised transmission intensity reduces the average age at infection²⁷.

Epidemiological parameters

To estimate longitudinal trends in the incidence of infection (as opposed to disease, when clinical signs are exhibited), a series of key epidemiological parameters must be determined and their relationship with demographic variables ascertained. We exclude consideration of the influences of host genotype and changes in the nature of the infectious agent over time. In rodent models,

FIG. 1 *a*, Longitudinal trend in the incidence of confirmed BSE cases per half year from 1986 to the first half of 1996. Note that the number for the first half of 1996 will rise once reporting delays are taken into account (see *b*). *b*, Distribution of the reporting delay from diagnosis to entry on the Central Veterinary Laboratory BSE database for confirmed cases (mean, 3.1 months). *c*, Number of cattle (all ages) and calves up to 12 months old in Great Britain from 1974 to 1995 (data from MAFF^{49,50} and The Scottish Office⁵¹⁻⁵³ for 30 June each year; these data exclude small holdings). Numbers of calves are used as inputs for the demographic model. *d*, Estimated survivorship function for cattle showing probabilities of survival to each age, $S(u)$, estimated using standard methods⁵⁴ with data on age structure for 1982, 1988, 1989, 1991 and 1994 (*e*) and correcting for declining herd size by scaling by the size of the birth cohort (data from MAFF^{49,50}, The Scottish Office⁵¹⁻⁵³ and Milk Marketing Board⁵⁵, assuming average inter-calf interval of 12.5 months⁵⁶). Estimated $S(u)$ values do not vary significantly through time, and geometric mean values (shown here) are used throughout. *e*, Estimated age structure of the British cattle herd in 1994. Note the substantial loss through slaughter between ages 2 and 3 years. Data are the same as in *d*. *f*, Kaplan-Meier curves based on unpublished research on the duration of the incubation period of BSE in cattle dosed orally with a range of varying quantities of brain from animals exhibiting clinical signs (sample sizes: 10, 1g; 9, 10g; 10, 100g; 10, 3 × 100g). *g*, Spatial distribution of the incidence of BSE per head of cattle from 1986–95 by county.



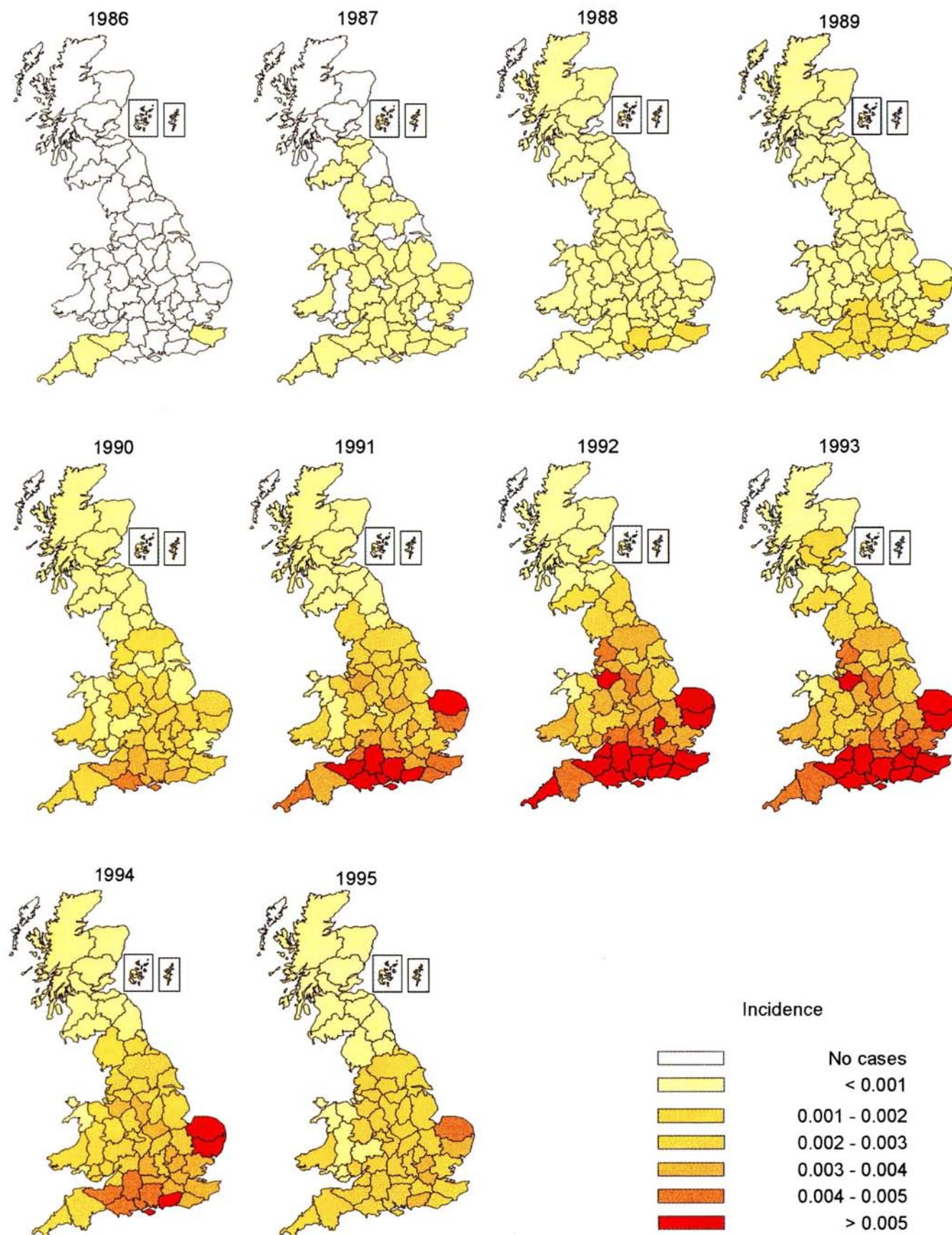
where scrapie prions passed in one species are inoculated into another, recycling of infection can act to reduce the incubation period of the disease^{24,28}.

Age-dependent exposure and susceptibility. Changes in the incidence of infection over time, and by host age, depend on the age-dependent rates of exposure and susceptibility (together given by probability density function $g(a)$), weighted by time-dependent levels of exposure (dose of aetiological agent), $K(t)$. The latter is influenced by recycling of BSE-infected tissues in MBM before 1988 and the degree to which the feed ban was effective after 1988. The age-dependent rate of exposure reflects the age-dependent intake of supplementary feed containing MBM. Supplementary feed is used in both dairy and beef herds, although management practices vary considerably between the two (over 90% of BSE cases originate in dairy or mixed herds). In herds where supplementary feed is given to calves, intake rises with age for the first two years, although it varies according to season and is highest during the winter. In many herds, intake is highest in cows over two years old. However, the epidemiological data on clinical onset

by age, taking account of the long incubation period of BSE, suggest that most, but not all, infection occurs during the first two years of life. Thus there may be age-dependent susceptibility to infection. For TSEs in rodent models, some experimental data hint at decreasing susceptibility with age²⁹, although other data show an opposite effect³⁰; experimental evidence for BSE in cattle is not yet available. The major changes in the immunology and physiology of the intestinal tract that occur in cattle as they age^{31,32} may well lead to (perhaps complex) age-related changes in susceptibility.

Incubation period. Interim results from recent experimental studies involving the oral dosing of cattle of known age with varying doses of brain from cattle affected with BSE (1g, 10g, 100g and 3 × 100g) (Fig. 1f) point to a dose-dependent incubation period, with a relatively small variance which also depends on dose. Across all doses, a range of 3–5 years is apparent, which will be increased once the surviving animals with longer incubation periods are included. The degrees to which animals at different stages of the incubation period are infectious to cattle through contaminated feed, to offspring born to infected dams, and to

g



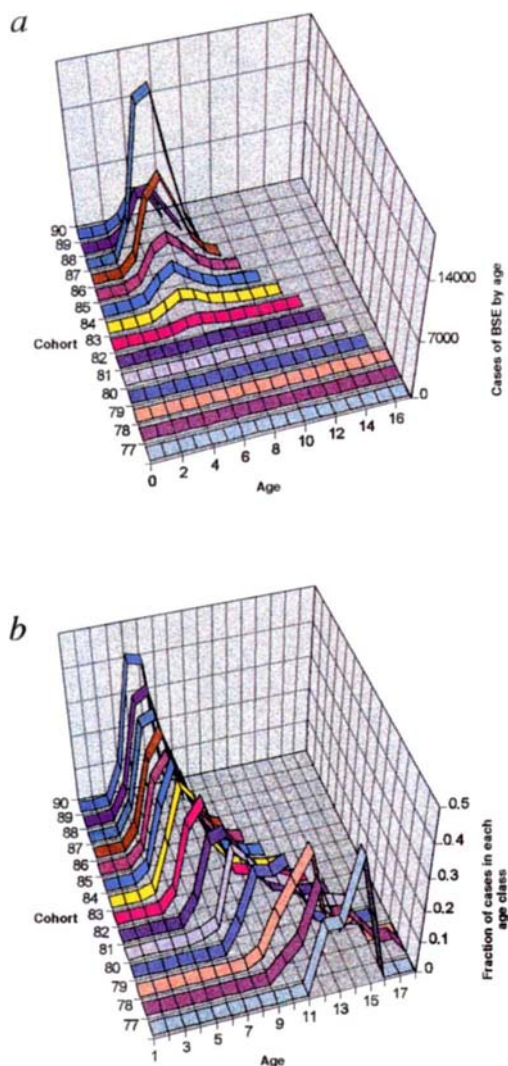


FIG. 2 a, Age distribution of confirmed cases of BSE by birth year cohort, b, Fraction of BSE cases in each age class for birth cohorts 1977–1990.

humans through contaminated food are key questions. Unlike scrapie in mice, where both spleen and brain are infectious, detectable titres of abnormal prion seem to be limited to brain, cervical spinal cord, terminal spinal cord, retina and possibly distal ileum in cattle^{17,33,34}. Experimental studies of TSEs in animal models, however, suggest the possibility, that infectivity will change markedly over the incubation period. Detectable titres of replicating abnormal prion in the brain might rise exponentially in the second half of the incubation period (these plateau much earlier in spleen and lymph nodes^{5,35}, and may decline subsequently³⁶), with very high levels after clinical signs have developed³⁷. For BSE in cattle, therefore, high infectivity may be restricted to the latter stage of the incubation period with a peak when clinical signs of disease appear.

Demography of cattle in Great Britain. The number of cattle in Great Britain has steadily fallen from 13.6 million in 1974 to 9.5 million in 1995 (Fig. 1c), and will decrease further as a result of new culling measures to control BSE. Recruitment has also declined during this period (Fig. 1c). From the age structure of the herd in different years it is possible to estimate the age-specific survivorship function (Fig. 1d), which gives an estimated life expectancy at birth of 3.0 years. Overall, these demographic patterns have caused a small increase in the mean age of the

national herd during the BSE epidemic. Between the ages of 2 and 3 years, a large fraction of animals are slaughtered for consumption (Fig. 1e). Those not slaughtered at that time are kept for breeding and/or milk production until, on average, they are 5–6 years old. Almost all cattle slaughtered (including those from dairy herds) provided beef and beef products for human consumption until the recent ban on cattle aged over 30 months. Age-specific slaughter rates are essential to the interpretation of the BSE epidemic and any associated risks to humans of infected animals entering the human food chain. Given an incubation period of 5 years, an average age of infection of about 1 year, and a life expectancy at birth of 3 years, it can be seen that most animals acquiring infection do not survive long enough for BSE to develop. Given that infectivity of tissues from animals infected with the BSE agent may depend on time since infection, any risk assessment for humans must include the distribution of slaughtered infected animals by time since infection and any changes therein over the course of the epidemic.

Maternal and horizontal transmission. To date no evidence exists to support the notion that the BSE agent can be transmitted horizontally either through close contact between susceptible and infected animals or through contaminated pasture¹⁶. This route cannot be eliminated, given evidence for the clustering of cases in individual herds, but the overall epidemiological pattern of infection and the pattern of decay in the epidemic after the feed ban argue against its occurrence at a level high enough to influence materially the course of the epidemic. Maternal transmission is thought to occur in scrapie³⁸, and horizontal transmission must also occur owing to the long-term persistence of the disease, and so this relevance to the interpretation of the British BSE epidemic has long been recognized⁴. A recent cohort study that monitored calves born to infected dams provided clear evidence of maternal transmission with a probability of 10.6% and a 95% confidence interval of 5.7–15.6 ($n = 546$)⁶. The infection rate for calves born to dams over the entire incubation period (5 years on average) may be lower (for example, if the rate is 10% over the last half-year of incubation and close to zero before that time, the average rate over a five-year incubation period is 1% if the birth rate is constant). The results of the cohort study provide no evidence that the incubation period following maternal transmission is different to that arising following feed-borne infection (the mean period from the maternal study is 5 years; ref. 6).

Trends in the incidence of infection

In the absence of an *in vivo* diagnostic test to detect infection, back-calculation methods using observed trends in the incidence of disease, and a knowledge of the distribution of the incubation period^{39,40}, provide the only methods available to estimate trends in the incidence of new infections. They should take into account the demography of the host species (age-specific survivorship of cattle and changes in recruitment over time), the maternal and feed-borne transmission routes, and the age-dependent exposure/susceptibility to infection. Simple projection methods can be used to predict future trends in cases of BSE^{25,41}, based on past trends in yearly cohorts, but this approach does not provide information on the pattern of infection, which is crucial to the full assessment of the current and future pattern of the epidemic. Our analyses are based on a model of the following structure, where $c(u, t_0)$ defines the probability density function (PDF) for the onset of BSE at age u for an animal born at time t_0 :

$$c(u, t_0) = \rho(t_0 + u)S(u) \times \left\{ [1 - \pi(t_0)] \int_{a=0}^u K(t_0 + a)g(a)f(u - a)da + \pi(t_0)f(u) \right\}$$

where $\pi(t_0)$ is the probability that an animal born at time t_0 is maternally infected, $S(u)$ is the age-dependent cattle survivorship function, $g(a)$ is the PDF for age at infection, $K(t)$ is the time-

dependent feed-risk function, $f(i)$ is the PDF for the incubation period, and $\rho(t)$ is the probability that a case diagnosed at time t was reported (to take account of under-reporting) (see legend to Fig. 3 for more detail). The probability $\pi(t_0)$ can be expressed as

$$\pi(t_0) = \left\{ \left[\sum_{n=1}^{\infty} \varepsilon^{n-1} (F - G)^{n-1} \right] \cdot \varepsilon G \cdot 1 \middle| t_0 \right\}$$

where ε is the probability that a calf born to an infected dam is itself infected, F and G (see legend to Fig. 3) are two operators used to calculate maternal infection from a dam infected by feed and maternal infection by a dam herself maternally infected, respectively, and n is the generation number for successive rounds of maternal infection. (In the time scale of the epidemic and for $\varepsilon \ll 1$, the effect of second- or higher-order generation terms in the maternal transmission function are likely to be negligible). Maximum-likelihood methods are used to estimate the variance of the incubation period (but not the mean), the age-dependent infection rate, and the time-dependent feed-risk func-

tion (see legends to Figs 3 and 4). Unfortunately, many parameters must be estimated in connection with the likelihood of feed-borne infection, using the age-structured notification records over time (data arise from a multinomial distribution when viewed by cohort and divided into yearly age classes). Thus the likelihood was maximized for various assumptions about the functional forms of $K(t)$, $g(a)$ and $f(i)$, given information on $S(u)$, ε and the mean of $f(i)$, using direction set^{42,43} and Latin hypercube sampling⁴⁴ along with simulated annealing methods⁴⁵⁻⁴⁷. A wide range of functional forms was examined for the age at infection distribution, Weibull and gamma distributions were considered for the incubation period, and the feed-risk profile was fitted using linear splines with optimized knot locations. Based on the minimization of the likelihood ratio χ^2 for all yearly cohorts, the model including 10% maternal transmission ($\varepsilon = 0.1$) for the last half-year of the incubation period, with the functional form for $f(i)$ and $g(a)$ shown in Fig. 3c, d, provided the best fit ($\chi^2_{219} = 482$) to observed trends of BSE cases in yearly cohorts of animals (illustrated in Fig. 4b). The fit to the observed changes in the age distribution of cases through time is consistent with there being no

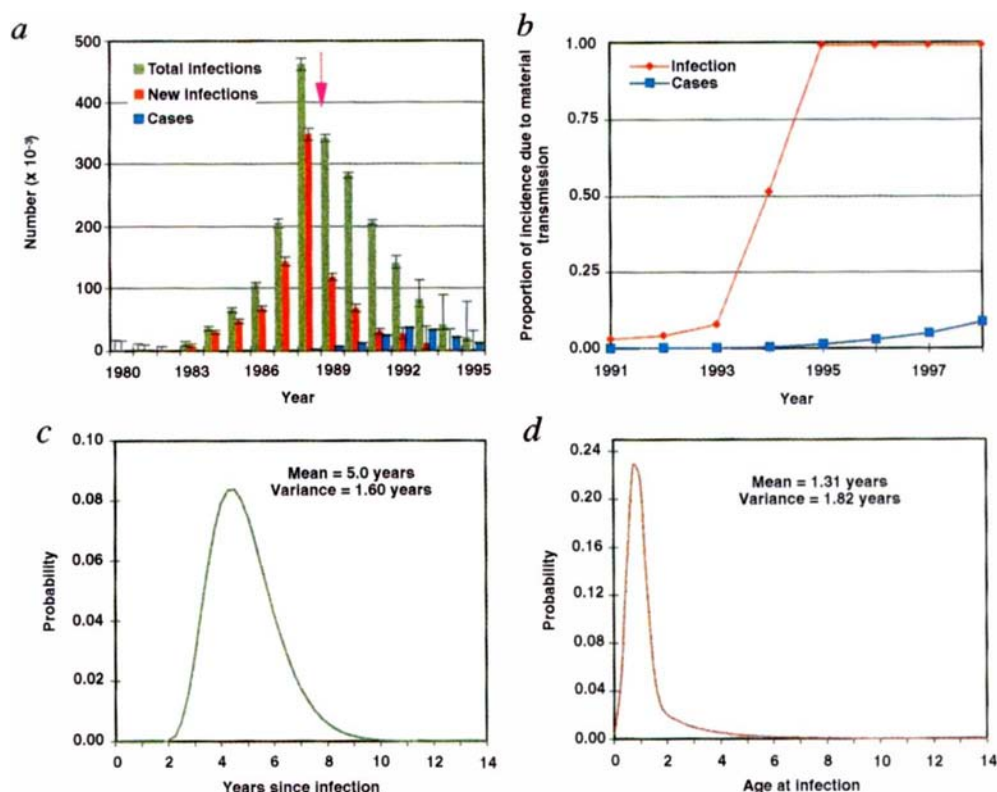


FIG. 3 a, Estimated trends in the incidence of infection, disease (per year) and the total number of infected cattle at the year end for Great Britain. The 95% confidence intervals (shown) were calculated from the hypercube obtained from the bounds of the univariate likelihood ratio confidence intervals, under the assumption that feed risk drops to zero by July 1996. (The arrow indicates the year of the statutory ban of the feeding of ruminant-derived protein to ruminants and the introduction of legislation making BSE a notifiable disease, with compulsory destruction of all animals exhibiting signs¹¹.) b, The proportions of infections and cases of BSE that are predicted to arise from maternal transmission (a rate of 10% over the last half year of the incubation period). The operators F and G used in the definition of $\pi(t_0)$ (see text) are defined as follows (y being any univariate function, $y(x)$):

$$(F \cdot y | t_0) = \left\{ \frac{\int_{x=0}^{z-2} B(x) S(z-x) \left[\int_{i=z-x}^{z-x+T} f(i) di \right] y(x) dx}{\int_{x=0}^{z-2} B(x) S(z-x) dx} \middle| z = t_0 \right\}$$

$B(x)$ is the birth rate of cows at time x , and T is the length of time before case diagnosis for which cows are assumed to be infectious to offspring (assumed to be 6 months or 1 year). Other quantities are defined elsewhere in the text. c, The estimated probability density function (PDF) $f(i)$, of the incubation period of BSE (gamma form). d, The estimated PDF of the age at infection distribution, $g(a)$. The cumulative distribution function has the form $g(a) = [1 - \exp(-(a/\alpha_1)^\beta)] [1 - \exp(-(a/\alpha_2)^{\beta+\gamma})]$.

$$(G \cdot y | t_0) = \left\{ \frac{\int_{x=0}^{z-2} B(x) S(z-x) \int_{a=0}^{z-x} K(x+a) g(a) \left[\int_{i=z-x-a+T}^{z-x-a} f(i) di \right] y(x) da dx}{\int_{x=0}^{z-2} B(x) S(z-x) dx} \middle| z = t_0 \right\}$$

change in the mean incubation period *per se*. Rather, these changes are in line with standard epidemiological theory: as the risk of infection increased up to 1989, the average age of infection declined and, as the risk of infection decreased after 1989, the average age of infection increased²⁷. Bearing in mind the problems raised by estimating many parameters from a single age- and time-structured data set, the goodness of fit provides a degree of confidence in the resulting incidence pattern (Fig. 3a). A large number of situations were examined, including: one in which maternal transmission is assumed to occur at a rate of 10% over the last year of the incubation period ($\chi^2_{219} = 492$); the best-fit case, in which a 10% maternal transmission rate is restricted to the last half-year ($\chi^2_{219} = 482$); and one in which maternal transmission is zero, with all transmission resulting from the consumption of contaminated feed ($\chi^2_{219} = 492$). These χ^2 values rise signifi-

cantly, regardless of model, if underreporting is excluded; when included, the estimated number of cases not reported before July 1988 is 3,240. The predicted pattern of maternal infection in recent years is shown as a proportion of total cases and infections in Fig. 3b. Examining predictions on a finer timescale (less than half-yearly) demonstrates evidence of seasonality in contaminated feed intake (Fig. 4c). If the period over which maternal transmission takes place is increased, the goodness of fit of the model decreases (for example, $\chi^2_{219} = 740$, for transmission allowed for 3 years before BSE in the dam).

Back-calculations of trends of infection incidence suggest that the feed ban introduced in mid 1988 had an immediate and lasting impact. Infection is predicted to have continued through this route, although at a much reduced level, before falling to negligible levels by mid 1994, after which new infections arose entirely

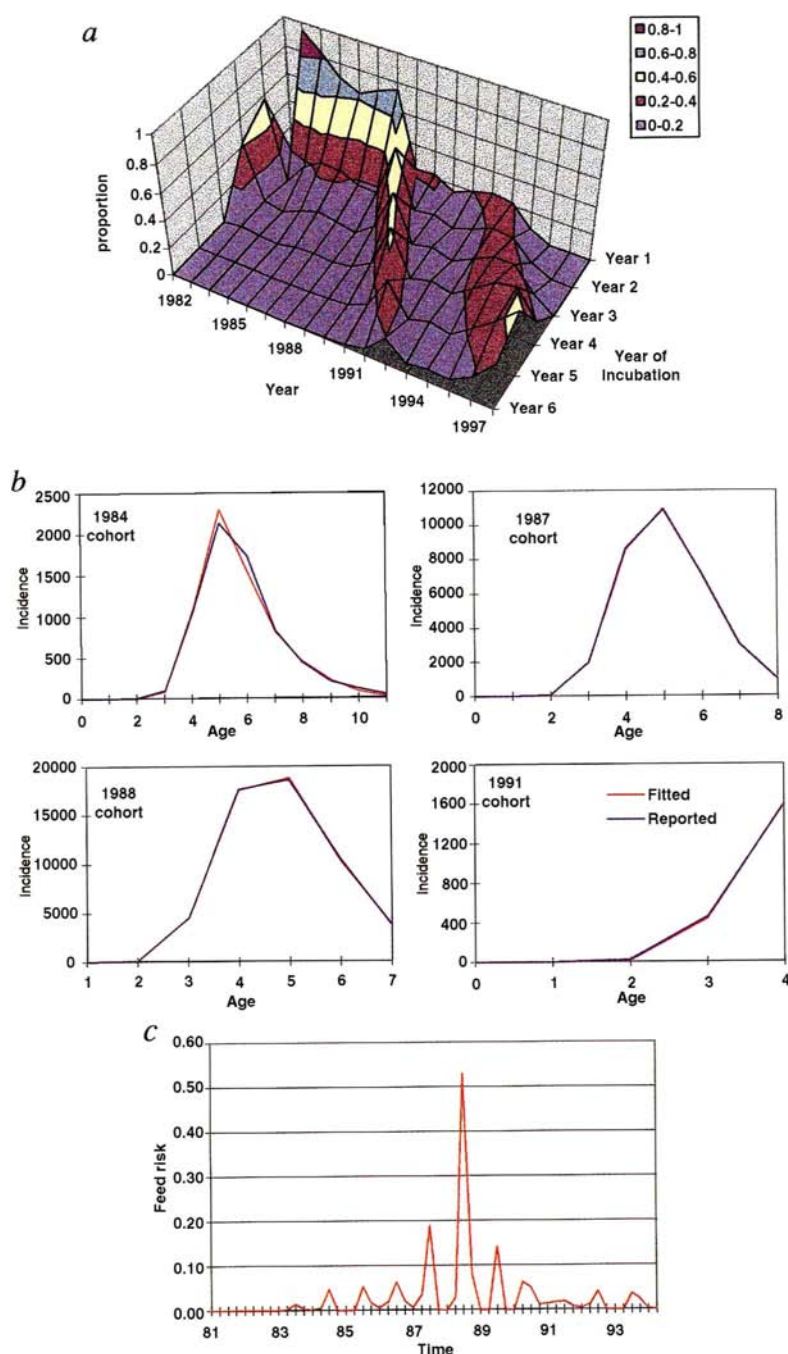


FIG. 4 a, The predicted proportion of infected animals in each age class by birth-year cohort using the model with 10% maternal transmission in the last half year of the incubation period of the dam. Although the proportion of older infected animals increases in recent years, the number of infected animals declines over the plotted time period. b, The observed and expected number of cases of BSE by age in a series of birth cohorts for the model with 10% maternal transmission in the last half year of the maternal incubation period (overall $\chi^2_{219} = 482$). c, The estimated time series of the risk of infection through contaminated feed over time ($K(t)$), illustrating the high risk during the winter months and the effect of the 1988 feed ban.

from the maternal route (318–352 in 1995) (Fig. 3*b*). The large difference between numbers infected and numbers of cattle diagnosed with BSE reflects the short life expectancy of cattle relative to the average incubation period, such that most infected animals were slaughtered before exhibiting disease. We estimate that approximately 446,000 (440,000–580,000) infected animals entered the human food chain before the specified bovine offal ban at the end of 1989 (ref. 48), with approximately 283,000 (270,000–330,000) more before the end of 1995. The estimated total number of animals infected over the period 1974 to the end of 1995 is 903,000 (840,000–1,250,000). Of these the estimated number of infections arising through maternal transmission is 5,100 (4,970–6,250) to the end of 1995, and 340 (240–1,270) in the period 1996–2001. It may take many years before an accurate assessment can be made of whether or not there is an epidemic of an apparently new variant of CJD, but if the postulated association between the new variant of CJD in humans and BSE is proved correct², then the numbers of infected animals that are culled and have entered the food chain, and the distribution of such numbers by stages of the incubation period in infected cattle (associated with a possible higher risk later in the incubation period), are of considerable public health significance. Maternal transmission is not predicted to lengthen or increase the scale of the epidemic (Fig. 5).

In the rising stage of the epidemic, most culled infected animals were in the early stages of incubation, although as the epidemic moves into the declining phase a growing fraction are in the late stages (note, however, that the total number of cases is declining at this stage) (Fig. 4*a*). At first sight the predictions plotted in Fig. 4*a* might be interpreted as the risk to humans being greater in recent years (1991–1995) than at the peak of the epidemic of infection (1987–1990). This is unlikely to be the case because, before August 1988, animals showing clinical signs (and hence with very high levels in the CNS of the BSE agent) could enter the food chain. The specified offal ban came into force in 1989, and over the following years infected brain and spinal cord should have been eliminated from the human food chain.

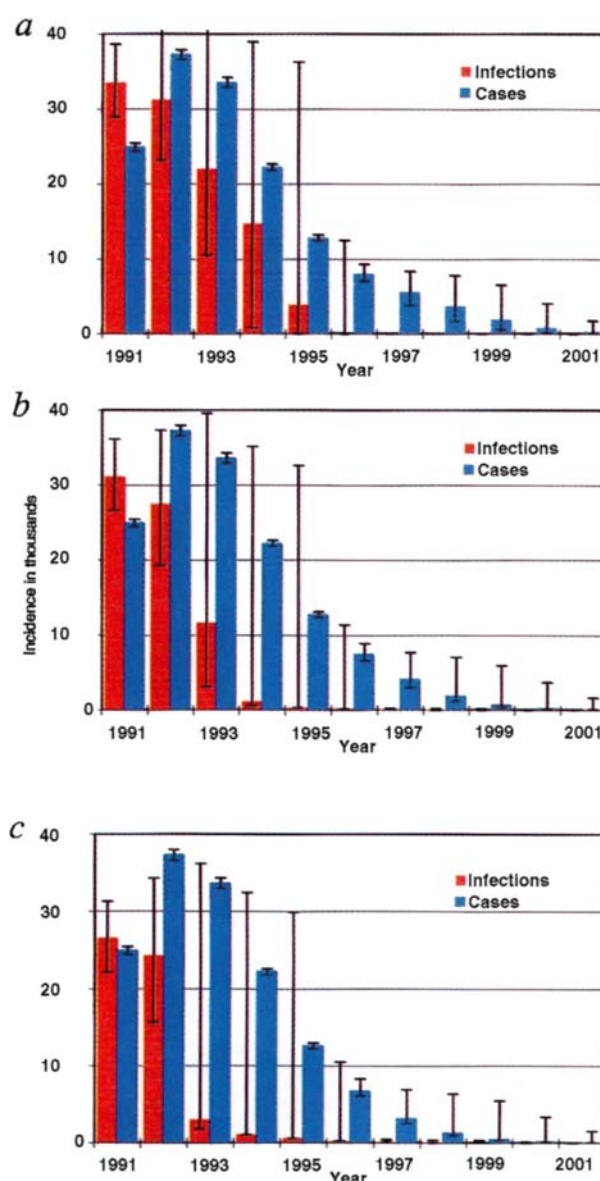


FIG. 5 Predictions of future trends in the incidence (per year) of cases of BSE and infections for the model with no maternal transmission (a), 10% maternal transmission in the last half year of the maternal incubation period (b), and 10% maternal transmission in the last year of the maternal incubation period (c). In fitting the trend to data on cases before 1996, the best fit was achieved by the model *b*. Error bars denote 95% prediction intervals calculated as described in Fig. 3.

TABLE 1 Predicted trends in the incidence (by year) of BSE infections (a) and cases (b) from 1996 to 2001 under different assumptions

(a)	No maternal transmission		6 months, 10% maternal trans.		1 year, 10% maternal trans.	
	Expected value	95% Prediction interval	Expected value	95% Prediction interval	Expected value	95% Prediction interval
1996	0	(0–12,500)	189	(155–11,300)	270	(224–10,500)
1997	0	(0–0)	95	(63–236)	102	(75–435)
1998	0	(0–0)	38	(21–214)	31	(20–389)
1999	0	(0–0)	12	(5–162)	8	(5–276)
2000	0	(0–0)	3	(1–86)	2	(1–136)
2001	0	(0–0)	1	(0–33)	0	(0–49)
(b)	No maternal transmission		6 months, 10% maternal trans.		1 year, 10% maternal trans.	
	Expected value	95% Prediction interval	Expected value	95% Prediction interval	Expected value	95% Prediction interval
1996	7,988	(7,044–9,306)	7,386	(6,541–8,856)	6,740	(6,085–8,291)
1997	5,573	(3,774–8,369)	4,111	(3,006–7,664)	3,145	(2,559–6,904)
1998	3,644	(1,654–7,762)	1,864	(1,153–7,025)	1,247	(960–6,365)
1999	1,896	(560–6,545)	682	(388–5,909)	456	(357–5,417)
2000	744	(152–4,042)	221	(128–3,660)	172	(140–3,388)
2001	225	(34–1,728)	72	(45–1,592)	68	(56–1,499)

Note that for the case of 10% maternal transmission (trans.) over 6 months, the predicted number of cases in the period 1997–2001 is 6,950. See text and legends to Figs 3 and 5.

The future

Extrapolation of the longitudinal trends in incidence of new infections (ignoring for now the impact of culling) from 1996 to 2001 gives the pattern of infections and cases of disease detailed in Table 1. This includes the appropriate prediction intervals for models with no maternal transmission but continued risk from feed, contaminated feed plus 10% maternal transmission rate for the last half-year of maternal incubation, and contaminated feed plus 10% maternal transmission for the last year of maternal incubation. In the last two models the risk from contaminated feed is estimated to have fallen to zero by mid 1994; it is likely that all of the new infections since then have arisen through maternal transmission (Fig. 3b).

In analyses to discriminate between different culling policies, optimization is based on the criteria of maximizing the number of BSE cases prevented and minimizing both the number of animals culled and the number of herds involved in the cull. In such calculations, several factors are important: the number of infected animals when the policy is implemented; the age distribution of infected animals; and the distribution of infection between herds. The first two factors can be assessed from the output of the models (Figs 5 and 6a, b). The predicted age distributions provide a means of assessing the expected impact of a cohort-targeted culling policy. Similarly, by distinguishing cases arising through feed or through maternal transmission (Fig. 3b), the expected impact of a maternally targeted policy can be assessed. A disease-case-herd-targeted policy involves identification of BSE cases over a defined period, and slaughtering cattle born on the same farms. A variant

of this policy is to define an incidence of BSE (number of cases in a defined cohort per head of cattle in the holding) above which the cohort is culled; this is a herd-incidence-targeted policy.

The likely success of herd-targeted policies depends on the degree to which future cases are clustered in cohorts within herds where animals diagnosed with BSE have originated. There is an important distinction to be made here between the distribution of infected animals and the distribution of diagnosed cases of BSE. In roughly one-third of those herds in which BSE-diagnosed cattle were born, no cases have been reported resulting from movement of incubating cattle from those herds. Whether or not cattle movements are related to the detection of early signs of disease (before firm diagnosis) is difficult to assess, but the distribution of the time interval between when an animal is moved between herds and the date of diagnosis shows an unusual bimodal pattern with an early peak. This may suggest that perhaps 1% of the total animals with BSE had been moved between herds close to the point of disease diagnosis. However, animal movement may result in a stress-induced acceleration of the appearance of clinical signs. Another important factor in cohort-based herd-targeted culling (on the basis of diagnosed cases of disease) is the clustering of cases by herd size. If the distribution is clustered within all herd size categories then targeting is more likely to be successful in removing infected (but not yet diseased) animals. There is clustering, with the distribution of case numbers in each herd size category (<50 to >200) being overdispersed with variances significantly greater than the respective means. This acts to enhance the likelihood of removing infected animals by a disease-herd-

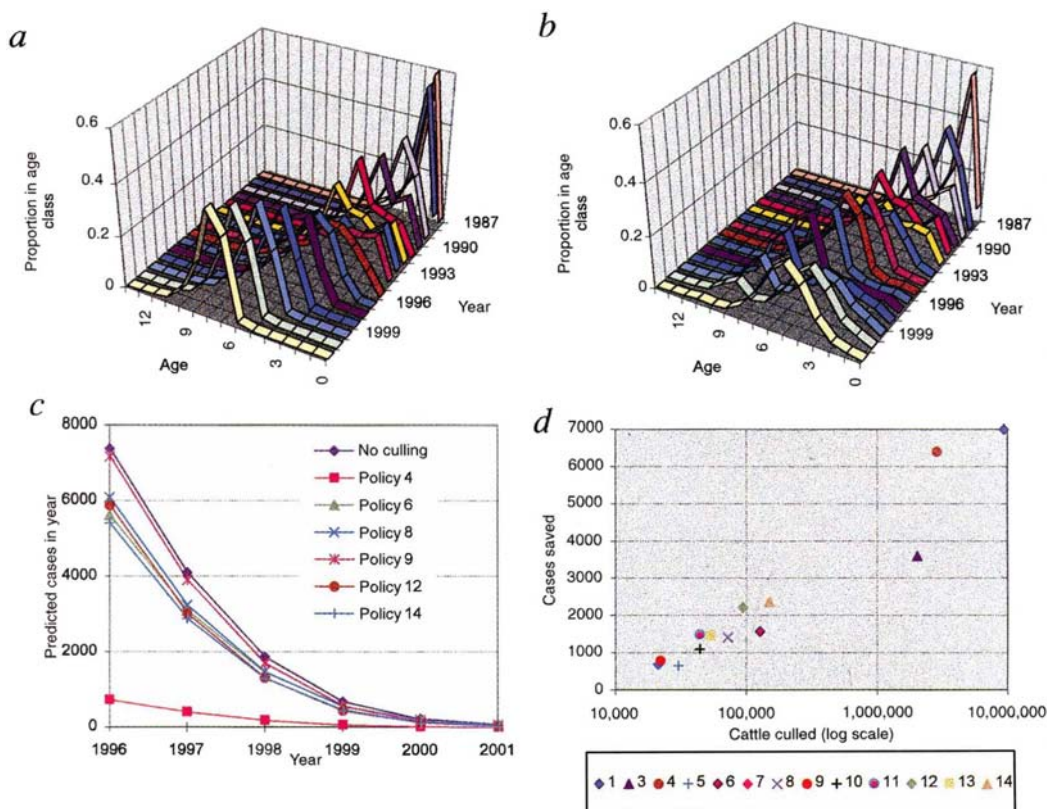


FIG. 6 a, b, The predicted age distribution (proportion in age class) of infected animals after 1995 for two of the three models described in Fig. 5. a, No maternal transmission; b, 10% maternal transmission in the last six months of the maternal incubation period. Note that proportions are plotted; the numbers of cases are small as detailed in Table 1. c, Predicted effect of selected culling policies on the future incidence of BSE cases, 1997–2001 (policies numbered as in Table 2). d, Comparison of the number of BSE cases saved and the number of cattle slaughtered (log

scale) for the different culling policies listed in Table 2. This comparison is relatively robust to uncertainty as to the number of future cases, but is sensitive to the estimated distribution of those cases in different herds and age classes. The overall pattern is one of diminishing returns across all culling policies, because, as the number of cattle slaughtered increases the number of cases increases much less rapidly. The numbers identify the policies detailed in Table 2.

targeted policy, over that which would pertain if the distribution was random. For herd-targeted policies, calculation of the fraction of cases captured by a defined policy cannot be precise owing to uncertainties in the future degree of clustering. The analyses are based on the calculation of the fraction of cases in the targeted age range captured by a given policy in 1995, 1994 and 1993, as if the policy had been implemented in those years, and extrapolating forwards. Estimates of the number of cases saved are therefore approximate. Most emphasis should be placed on the ranking of the relative efficiencies of different policies given that estimates of the number of holdings affected and cattle culled are fairly precise.

Working on the basis of a maternal transmission rate of 10% for the last half-year (the best-fit model to the data), predicted

impacts of a range of possible culling policies are compared in Table 2. One measure of the efficiency of any given policy is provided by the ratio of the number of animals culled to the predicted number of cases saved over the period from 1997 to 2001. This ratio must be considered in parallel with the number of herds affected by the cull and the total number of BSE cases (Table 2). The predicted total number of cases yet to be diagnosed in the absence of culling is 6,950 (see Table 1 for prediction intervals). Herd-targeted policies are by this measure significantly better than other options (Table 2), with a herd-incidence-targeted policy predicted to be most 'efficient'. For the latter approach, a threshold incidence of 1 case in the July 1989 to June 1992 cohort, per 27–50 cattle in the holding as a whole is most efficient and also affects the smallest number of herds. In light of

TABLE 2 Comparison of possible culling policies

No.	Culling policy description	Cases saved		Total cattle culled		No. of origin holdings	Cattle culled per case saved
		Number	%	Number	%		
	Non-targeted						
1	All cattle	6,950	100	9,360,000	100	111,000	1,300
	Age-targeted						
2	All cattle born before 7/88	250	4	352,000	3.8	≤ 111,000	1,400
3	All cattle born 10/90–6/93	3,600	51	2,030,000	22	≤ 111,000	564
	Herd-targeted (case)*						
4	All cattle born in herds from which a case originated during 1/91–12/95	6,300	90	2,870,000	31	28,500	455
5	Cattle born in the 10/90–6/91, 7/91–6/92 or 7/92–6/93 cohorts in herds from which a case originated in the corresponding cohort during 1/91–12/95 (govt policy)	650	9	30,100†	0.32	1,460	46
6	As 15), but extended to include 7/89–9/90 cohort (govt compulsory + voluntary policy).	1,580	23	127,000†	1.4	6,240	80
	Herd-targeted (incidence)*‡						
7	Cattle born in 7/89–6/92 in herds with more than 1 case in that cohort range (during 1/91–12/95) per 27 cattle in the holding as a whole.	691	10	21,300†	0.23	638	31
8	As 7) but with a threshold of 1 case per 50 cattle	1,420	20	71,900†	0.77	2,000	51
	Maternally targeted§						
9	Cattle born after 10/90 within six months of BSE case in the dam	797	11	<22,000	0.24	≤ 22,000	28
10	Cattle born after 10/90 within 12 months of BSE case in the dam	1,100¶	22	<44,000	0.47	≤ 35,600	40
	Combined herd-targeted and maternally targeted policies						
11	* Incidence (1 per 27) + § maternally targeted policy = policies 7 and 9 combined	1,490	21	<44,000†	0.47	≤ 25,500	30
12	* Incidence (1 per 50) + § maternally targeted policy = policies 8 and 9 combined	2,220	32	<94,000†	1.0	≤ 26,800	42
13	* Govt compulsory + maternally targeted policy = policies 5 and 9 combined	1,450	21	<53,000†	0.57	≤ 26,300	37
14	* Govt inc. voluntary + maternally targeted policy = policies 6 and 9 combined	2,380	34	<150,000†	1.6	≤ 31,000	63

Projections are made using the model with 10% maternal transmission over 6 months (unless otherwise stated) and refer to cases saved over the period 1997 to 2001 achieved by a culling policy implemented at the end of 1996. Values are to 3 sig. figs, percentages to 2 sig. figs.

* These projections require estimates of the fraction of cases captured by a diseased-herd-targeted policy. These estimates were obtained by calculating the appropriate fraction for the years 1995, 1994, and 1993 (had the policy been implemented in those years) and extrapolating to 1996. In cases where the exact date of birth of a case was not listed in the database, this was calculated from the estimated age.

† Using age structure of BSE-infected herds calculated from lactation data collected by CVL.

‡ Incidence is calculated from cases in specified cohort range, but using the total number of all animals on each holding as estimated from the BSE case database, since reliable estimates of total numbers of animals in the targeted cohort on each holding are not available. For holdings with multiple herds, the estimated size is the sum of the sizes of the constituent herds that appear in the database, so if herds within a holding have not reported a case, they are not included in the estimate of holding size. Two representative policies are shown here, illustrating that by lowering the incidence threshold for culling, more cases are saved, but at the cost of disproportionately more animals culled.

§ Running policy (culling until 2001). Other policies involve one-off cull in 1996.

|| Assuming 6 month infectious period for maternal transmission.

¶ Assuming 10% maternal transmission occurs in the last 12 months of incubation period and calculated from corresponding model (predicting 5,100 cases total in 1997–2001).

recent evidence for maternal transmission⁶, consideration must be given to augmenting policies by culling cattle in given cohorts born to dams who subsequently developed BSE (a maternally targeted policy). Maternal targeting is very efficient and is predicted to reduce to almost zero the number of new infections arising over the period 1997 to 2001, but a large number of herds would be affected by the cull. Combining a herd-incidence-targeted policy with a maternally targeted policy is predicted to maximize efficiency.

Conclusions

The epidemic of BSE in British cattle seriously affected both animal health and Europe's agricultural industry. However, the epidemic is well past its peak, and seems to be in a phase of rapid decline. New infections from contaminated feed are predicted to be close to zero by the end of 1994, with all new cases of infection arising from maternal transmission. However, the numbers are small (Table 1) and this route of infection by itself cannot sustain the epidemic. To accelerate the rate of decline of the epidemic (Fig. 6c) without the slaughter of a very large number of animals

(Fig. 6d) is very difficult in the absence of an *in vivo* diagnostic test. Bearing in mind both this and the prediction that the epidemic is likely to fade close to extinction by the year 2001 in the absence of culling, the calculated optimal culling policy (maximizing cases saved, minimizing animals culled) is a combination of herd-incidence targeting (with a threshold for culling of one BSE case in cattle born July 1989 to June 1992 per 27 cattle in the holding as a whole) plus a maternally targeted policy. However, the practicalities of such an approach, given that roughly 25,500 herds would be affected, need careful consideration.

We believe eradication to be possible, but the history of the epidemic in cattle will continue to be of great importance as attempts are made to formalize risk assessments to determine the likely degree of past exposure of the human population to the BSE agent, and therefore the likelihood of an association between BSE and the apparently new variant of CJD². Whether or not the 12 reported cases of the new variant are the beginning of an epidemic remains uncertain, and will continue to be so for the next few years. □

Received 19 July; accepted 12 August 1996.

- Statement of 20 March 1996 by Secretary of State for Health Stephen Dorrell to House of Commons, London.
- Will, R. G. et al. *Lancet* **347**, 921–925 (1996).
- Prusiner, S. B., Telling, G., Cohen, F. E. & DeArmond, S. J. *Sem. Virol.* **7**, 159–173 (1996).
- Report of the Working Party on Bovine Spongiform Encephalopathy. (DoH, MAFF, London, 1989).
- Spongiform Encephalopathy Advisory Committee *Transmissible Spongiform Encephalopathies: A Summary of Present Knowledge and Research* (HMSO, London, 1995).
- Wilesmith, J. W., Wells, G. A. H., Ryan, J. B. M., Gavier-Widen, D. & Simmons, M. M. *Nature*, (submitted).
- Programme to Eradicate BSE in the United Kingdom* (HMSO, London, 1996).
- Middleton, D. J. & Barlow, R. M. *Vet. Rec.* **132**, 545–547 (1993).
- Foster, J. D., Hope, J. & Fraser, H. *Vet. Rec.* **133**, 339–341 (1993).
- Wilesmith, J. W., Ryan, J. B. M. & Atkinson, M. J. *Vet. Rec.* **128**, 199–203 (1991).
- The Bovine Spongiform Encephalopathy Order 1988*. Statutory Instrument 1988 No. 1039 (HMSO, London, 1988).
- Wilesmith, J. W., Ryan, J. B. M., Hueston, W. D. & Hoinville, L. J. *Vet. Rec.* **130**, 90–94 (1992).
- Wilesmith, J. W. & Ryan, J. B. M. *Vet. Rec.* **132**, 300–301 (1993).
- Kimberlin, R. H. & Wilesmith, J. W. *Ann. N.Y. Acad. Sci.* **724**, 210–220 (1994).
- Wilesmith, J. W. *Phil. Trans. R. Soc. Lond. B* **343**, 357–361 (1994).
- Hoinville, L. J., Wilesmith, J. W. & Richards, M. S. *Vet. Rec.* **136**, 312–318 (1995).
- Fraser, H., Bruce, M. E., Chree, A., McConnell, I. & Wells, G. A. H. *J. Gen. Virol.* **73**, 1891–1897 (1992).
- Hunter, N., Foster, J. D., Dickinson, A. G. & Hope, J. *Vet. Rec.* **124**, 364–366 (1989).
- Goldmann, W., Hunter, N., Smith, G., Foster, J. & Hope, J. *J. Gen. Virol.* **75**, 989–995 (1994).
- Foster, J. D., Bruce, N., McConnell, I., Chree, A. & Fraser, H. *Vet. Rec.* **138**, 546–548 (1996).
- Dawson, M., Wells, G. A. H., Parker, B. N. J. & Scott, A. C. in *Sub-acute Spongiform Encephalopathies. Current Topics in Veterinary Medicine and Animal Science* (eds Bradley, R., Savel, M. & Marchant, B. A.) 25–32 (Kluwer Academic, Dordrecht, The Netherlands, 1991).
- Martin, T. et al. in *Proc. European Soc. Veterinary Virology*, September 23–26 148 (Uppsala, Sweden, 1991).
- Hunter, N., Goldmann, W., Smith, G. & Hope, J. *Vet. Rec.* **135**, 400–403 (1994).
- Kimberlin, R. H. *Intervirology* **35**, 208–218 (1993).
- Stekel, D. J., Nowak, M. A. & Southwood, T. R. E. *Nature* **381**, 119 (1996).
- Wilesmith, J. W. in *Proceedings of the Vth International Workshop on Bovine Spongiform Encephalopathy: The BSE Dilemma* (ed. Gibbs, C. J.) (Serono Symposia, Norwell, USA).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).
- Weissman, C. *Nature* **349**, 569–571 (1991).
- McKinley, M. P., DeArmond, S. J., Torchia, M., Mobley, W. C. & Prusiner, S. B. *Neurology* **39**, 1319–1324 (1989).
- Outram, G. W., Dickinson, A. G. & Fraser, H. *Nature* **241**, 536–537 (1973).
- Hall, J. G. in *The Ruminant Immune System in Health and Disease* (ed. Morrison, W.) (Cambridge University Press, Cambridge, 1986).
- Madge, D. S. *The Mammalian Alimentary System: A Functional Approach* (Arnold, London, 1975).

- Wells, G. A. H. et al. *Vet. Rec.* **135**, 40–41 (1994).
- Ministry of Agriculture Fisheries and Food *Bovine Spongiform Encephalopathy in Great Britain: A Progress Report* (MAFF, 1996).
- Kimberlin, R. H. & Walker, C. A. *Ciba Foundation Symposium* **135**, 37–62 (1988).
- Asher, D. M., Gibbs, C. J. & Gajdusek, D. C. *Ann. Clin. Lab. Sci.* **6**, 84–103 (1976).
- Dickinson, A. G. & Outram, G. W. in *Slow Transmissible Diseases of the Nervous System* (eds Prusiner, S. B. & Hadlow, W. J.) 13–31 (Academic, New York, 1979).
- Dickinson, A. G., Stamp, J. T. & Renwick, C. C. *J. Comp. Pathol.* **84**, 19–25 (1974).
- Brookmeyer, R. & Gail, M. H. *J. Am. Stat. Assoc.* **83**, 301–308 (1988).
- Bacchetti, P., Segal, M. R. & Jewell, N. P. *Stat. Sci.* **8**, 82–101 (1993).
- Richards, M. S. et al. in *Proc. Soc. Veterinary Epidemiology and Preventative Medicine* (ed. Thushfield, M. V.) 70–81 (1993).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1992).
- Jacobs, D. A. H. *State of the Art in Numerical Analysis* (Academic, London, 1977).
- McKay, M. D., Beckman, R. J. & Conover, W. J. *Technometrics* **21**, 239–245 (1979).
- Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. *Science* **220**, 671–680 (1983).
- Kirkpatrick, S. *J. Stat. Phys.* **34**, 975–986 (1984).
- Vanderbilt, D. & Louie, S. G. *J. Comput. Phys.* **56**, 259–271 (1984).
- The Bovine Offal (Prohibition) Regulations 1989*. Statutory Instrument 1989 No. 2061 (HMSO, London, 1989).
- Ministry of Agriculture Fisheries and Food *Agricultural Statistics—United Kingdom 1974–1989* (HMSO, London, 1975–1990).
- Ministry of Agriculture Fisheries and Food *The Digest of Agricultural Census Statistics—United Kingdom 1991–1994* (HMSO, London, 1992–1995).
- Department of Agriculture and Fisheries for Scotland *Agricultural Statistics—Scotland 1974–1979* (HMSO, Edinburgh, 1975–1980).
- Department of Agriculture and Fisheries for Scotland *Economic Report on Scottish Agriculture 1980–1989* (HMSO, Edinburgh, 1981–1990).
- The Scottish Office *Economic Report on Scottish Agriculture 1990–1994* (HMSO, Edinburgh, 1991–1995).
- Caughley, G. *Analysis of Vertebrate Populations* (Wiley, Chichester, UK, 1980).
- Milk Marketing Board *National Milk Records, Monthly Statement of Milk Records* (Thames Ditton, Surrey, UK, 1991).
- Esslemont, R. J. & Kossabati, M. A. (DAISY, Department of Agriculture, University of Reading, UK, 1995).

ACKNOWLEDGEMENTS. R.M.A., C.A.D., N.M.F. and H.J.U. thank the Wellcome Trust for research grant support. M.E.J.W. thanks the Royal Society and the BBSRC. J.W.W., J.B.M.R. and L.J.H. are grateful for research funding from MAFF and for the technical and clerical support from other members of the BSE Research Team of the Epidemiology Department, Central Veterinary Laboratory, especially for the maintenance of the main epidemiological database. A.R.A. and G.A.H.W. thank MAFF for research funding and the technical support of the Pathology Department, CVL. We also thank D. Austin, C. Bostock, I. Morrison, T. Draycott and K. Parker for help.

CORRESPONDENCE and requests for materials should be addressed to R.M.A. (e-mail: roy.anderson@zoo.ox.ac.uk).

Wind-dominated optical line emission from accretion disks around luminous cataclysmic variable stars

N. Murray & J. Chiang

Canadian Institute for Theoretical Astrophysics, University of Toronto,
Toronto, Ontario M5S 3H8, Canada

CATAclysmic variable stars are thought to consist of a low-mass star orbiting a more massive white-dwarf primary star. Gas from the secondary star flows onto the white dwarf by way of an accretion disk, and the recurrent outbursts that characterize these objects are attributed to either instabilities in the accretion disk (in the case of dwarf novae) or thermonuclear explosions of material accumulated on the white dwarf (for novae). Observations^{1,2} of low-luminosity cataclysmic variables have revealed double-peaked emission lines consistent with the accretion-disk picture. But the emission lines associated with high-luminosity cataclysmic variables are single-peaked³⁻⁹, which poses a considerable challenge for the disk theory^{10,11}. It has been suggested that winds from the accretion disk, driven by radiation pressure, might be responsible for altering the emission-line characteristics^{4,10}. Here we present a model of the disk wind in which the emission lines arise in a thin layer where the wind emerges from the disk. In our model, the accelerating wind introduces a pronounced anisotropy in the optical depth of the disk, from which single-peaked emission lines naturally arise.

The observational case for accretion disks is moderately strong for cataclysmic variables (CVs), weaker for X-ray binaries and weakest for active galactic nuclei (AGN), including quasars. Several lines of evidence indicate the presence of a disk in CVs. First, the normally blue continuum turns red during eclipse in eclipsing cataclysmics, indicating hot material near the white dwarf and cooler material farther away¹². Second, in the recurrent outbursts in dwarf novae mentioned above, the rise in the ultraviolet flux lags in the optical range, which is consistent with mass transfer from the outer part of a disk to the inner part¹³. Third, double-peaked line profiles indicate the existence of receding and approaching material, as expected in a disk. The puzzle presented by single-peaked profiles in some luminous CVs (both dwarf novae in outburst and in nova-like variables, which look like dwarf nova that are stuck in the outburst state) appears to weaken the argument. This has led to claims that nova-like CVs do not have disks¹⁴.

The best evidence for accretion disks in AGN is the distribution and kinematics of mega-masers in a few nearby low-luminosity objects¹⁵. On the other hand, double-peaked profiles are rarely seen in AGN¹⁶. This suggests to some authors that AGN do not possess disks¹⁷. Motivated by the presence of blue-shifted absorption lines in some AGN spectra, Murray *et al.*¹⁸ have suggested that AGN possess line-driven winds emerging from the disk, and that radiation from the base of the wind produces the observed broad emission lines. The force generated by radiation pressure acting on lines in AGN disk winds was calculated at the same time as the ionization state of the wind. More recently, Murray and Chiang¹⁹ showed explicitly how radiative transfer effects in the base of the wind ensure that the profiles produced in AGN disks are single-peaked. Here we use the results of these AGN studies to explain why high-luminosity cataclysmic variables have single-peaked optical emission lines.

Like quasars, many high-luminosity CVs show blue-shifted absorption lines and this is direct evidence for outflowing winds. Early models of CV wind emission made the simplifying assumption

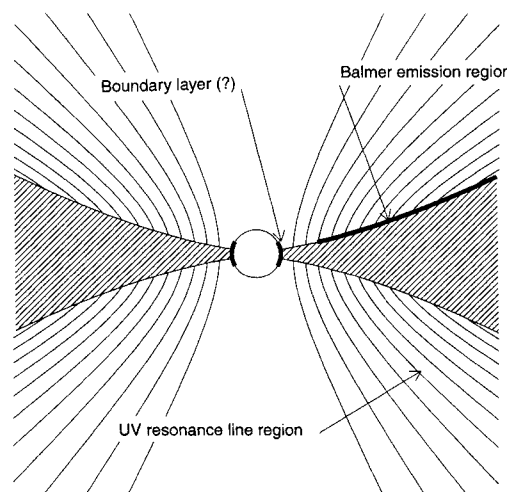


FIG. 1 A schematic cross-section through the star (in the centre of the figure and disk (shown hatched) in a nova-like cataclysmic variable, showing the wind and boundary layer. The boundary layer, which is expected theoretically but which seems not to be present in many objects, is depicted by the dark patches where the flow in the disk joins the surface of the white dwarf. The solid lines emerging from the disk represent wind streamlines. The ultraviolet (UV) lines form in an extended region in the wind as indicated. The single-peaked optical lines form where the wind emerges from the surface of the disk, shown by the dark patch labelled 'Balmer emission region'.

tion of radial outflow from the white dwarf^{20,21}. However, observations show that the resonance lines decrease in width during a partial eclipse. This is expected if the wind arises from the disk^{22,23}. Accordingly, we assume that the wind emerges from the disk. The geometry is shown in Fig. 1.

We have used the observed optical, ultraviolet and soft X-ray spectra^{24,25} of nova-like variables to calculate the ionization state, line force and dynamics of a CV disk wind. One complication is that no nova-like variable has been detected in the extreme ultraviolet (EUV). However, the spectra of dwarf novae in outburst are similar to those of nova-like variables, so we use the EUV spectra of SS Cygne²⁶ and VW Hydrae²⁷ instead. Following Murray *et al.* (ref. 18), we calculate the dynamics, ionization state (using the photoionization code CLOUDY; G. Ferland, personal communication) and line acceleration of a line-driven disk wind self-consistently. For a nova-like cataclysmic with a luminosity of 4×10^{34} erg s⁻¹ and a white dwarf mass of 0.5 solar masses, we find a wind mass loss rate of $\sim 10^{15}$ g s⁻¹ and a terminal velocity $v_\infty \approx 10,000 (10^9 \text{ cm}/r_i)^{1/2}$ km s⁻¹, where r_i is the radius at which a wind streamline emerges from the disk. The poloidal (non-azimuthal) velocity is given roughly by $v_p(r) = v_\infty(r_i)(1 - r_i/r)^{1/2}$, where r is the distance from the centre of the white dwarf and $\gamma \approx 1.4$ is a constant. Resonance ultraviolet lines such as C IV become optically thin at $v_p \approx 4,000$ km s⁻¹ in our models. The number density is $n \approx 7 \times 10^{13} \text{ cm}^{-3} (10^9 \text{ cm}/r)^2 (v_{th}/v_p)^2$, where $v_{th} \approx 2 \times 10^6$ cm s⁻¹ is the thermal velocity. We will see below that the ratio of the gradient of the poloidal velocity to the gradient of the azimuthal velocity, $(dv_p/dr)/(dv_\phi/dr)$ is an important quantity. From the expression for $v_p(r)$ we expect this ratio to be $\sim v_\infty/v_\phi \sim 3-5$. Numerically we find that it ranges from ~ 6.5 at the sonic point down to ~ 2 at the terminal velocity of the wind.

Having determined the properties of the wind, we turn to the calculation of the line profiles. Suppose we view the disk, assumed to lie in the x - y plane, from a point in the x - z plane at an inclination i . The monochromatic specific luminosity emitted by ions in such a disk is (see Horne²⁸; Murray and Chiang¹⁹)

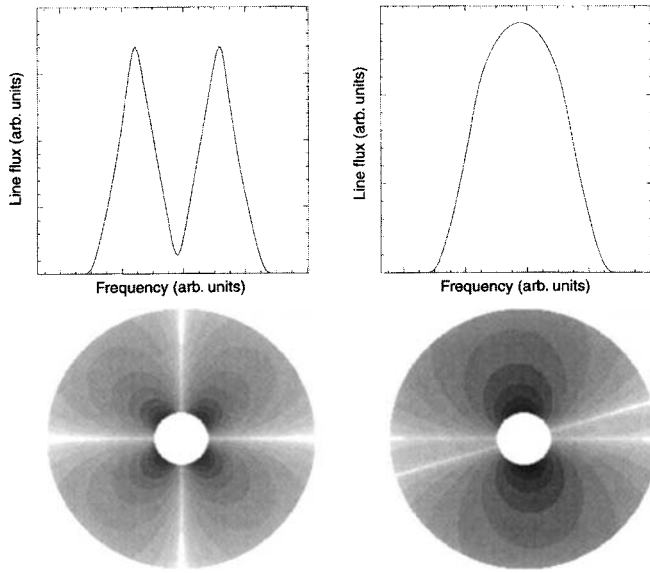


FIG. 2 Lower left, grey-scale plot of Q (see equation (4)) when there is no wind. The emission comes from four left-right and front-back symmetric lobes. The corresponding line profile (upper left) is double-peaked. Lower right, grey-scale plot of Q when there is wind with $(dv_p/dr)/(dv_\phi/dr) = 5$. The front and back of the disk, which have a low projected velocity, are much brighter than the left and right side. The resulting line (upper right) is single peaked. The disk inclination is $i = 87^\circ$ in both cases.

$$\mathcal{L}_v(\mathbf{n}) = \sum_{j=1,2} \int r dr S_v(r) \frac{l_{em}}{v_\phi} \frac{|Q_j| \sqrt{1 + (l_{sj}/l_{em})^2}}{\sin i |\cos \phi_j|} [1 - e^{-\tau_j}] \quad (1)$$

where the line optical depth is

$$\tau_j = \left(\frac{k_0 c}{\sqrt{\pi} v_0} \right) \frac{1}{|Q_j| \sqrt{1 + \left(\frac{l_{sj}}{l_{em}} \right)^2}} \quad (2)$$

The length of the vector from disk centre to the emitting point is r , and ϕ is the angle between this vector and the x -axis. $S_v(r)$ is the source function. The subscripts $j = 1, 2$ reflect the fact that at any radius there are two points on the disk where ions with a rest-frame transition at v_0 produce emission red- or blue-shifted to v . The quantity k_0 is the absorption coefficient, proportional to the number density of emitting ions and to the strength of the transition. The quantity $Q = \mathbf{n} \cdot \mathbf{A} \cdot \mathbf{n}$ is the double scalar product of the velocity strain tensor with the unit vector $\mathbf{n}$ in the direction of the line of sight. It measures the rate at which the velocity changes along the line of sight. If Q is large, the number of ions along $\mathbf{n}$ capable of absorbing line photons is small because ions a short distance away are red- or blue-shifted out of resonance. The Sobolev length is $l_{sj} = v_{th} \cos i / |Q_j|$, and the length scale of the emitting gas is l_{em} which is equal to the scale height of the disk atmosphere when there is no wind. If there is a line-driven wind emerging at an angle λ to the disc plane, then $l_{em} = \sin \lambda (v_{em}/v_\infty) r$, where v_{em} is the magnitude of the poloidal (non-azimuthal) velocity of the emitting gas and v_∞ is the terminal velocity of the wind. We expect emission when the poloidal velocity is low, that is, $v_{em} \approx v_{th} \ll v_\phi$, but as noted above, the gradient of the poloidal velocity is larger than the gradient of the keplerian velocity

$$\frac{dv_p}{dr} \gtrsim \frac{dv_\phi}{dr} \quad (3)$$

The significance of the last relation becomes apparent when we examine the expression for Q :

$$Q = \sin^2 i \left[\frac{dv_p}{dr} \cos^2 \phi + \frac{v_p}{r} \sin^2 \phi + (3/2) \frac{v_\phi}{r} \sin \phi \cos \phi \right] - \cos i \left[\frac{dv_p}{dr} (\sin i \cos \phi \cot \lambda + \cos i) + \frac{v_\phi}{(2 \sin \lambda r)} \sin i \sin \phi \right] \quad (4)$$

The surface brightness is proportional to Q for an optically thick line ($\tau \gg 1$); compare equation (1). Grey-scale plots of Q with and without a wind are shown with the corresponding line profiles in Fig. 2. In the plots dark regions correspond to strong emission. If there is no wind $Q \approx \sin \phi \cos \phi$, resulting in a four-lobed emission pattern and double-peaked line profiles. When there is a wind present, and the disk is viewed at large inclination, the $\sin^2 i \cos^2 \phi dv_p/dr$ term in Q ensures that the front and back of the disk emit more strongly than the left and right sides of the disk, assuming equation (3) is satisfied. Relative to a disk with no wind, more line photons escape along directions with low projected velocity, so the profile is single-peaked. At lower inclination (a situation not shown in Fig. 2) the $\cos i \sin i \sin \phi$ term tends to brighten the left and right side of the disk, and double-peaked profiles can result.

For this mechanism to explain the single-peaked Balmer lines seen in nova-like systems, the lines must be optically thick. In our CLOUDY calculations, the optical depth of H α runs from ~ 30 at the sonic point down to unity at $v \approx 10^7 \text{ cm s}^{-1}$; at the latter velocity the density is so low that we do not expect much emission. The fact that the emission is optically thick is consistent with the observed Balmer decrement in nova-like variables. The ratio of the number density of hydrogen in the first excited state (principal quantum number $n = 2$) to the total number density is typically 5×10^{-8} at the sonic point.

The ultraviolet resonance lines of luminous CVs form in the wind, and are also single-peaked but for a different reason. They are produced by scattering in an extended region of the wind, as indicated in Fig. 1. At these distances the rotational velocity of the wind is reduced well below the keplerian value in the disk where the wind arises, though it is still substantial. It is this reduction in v_ϕ rather than any anisotropic emission that produces the single-peaked ultraviolet lines. Observationally, the reduced rotation is demonstrated by the rather small red- and blueshifts of the line centres seen during eclipses. These shifts are much smaller than those of the optical lines either in our simulations of eclipses or in observations.

We expect this model to be applicable to the class of nova-like variables known as SW Sextantis stars³⁻⁹. As it stands, the model leaves unexplained one characteristic of SW Sextantis stars, namely absorption in the line cores near orbital phase 0.5 (when the primary is between us and the secondary). This absorption is likely to be due to the accretion stream (which feeds the disk) overflowing the initial impact at the outer edge of the disk and continuing to smaller radii, as suggested by Hellier and Robinson²⁹. They suggested that it was the overflow stream that distinguished SW Sextantis stars from other cataclysmic variables. However, their model does not produce single-peaked profiles at most orbital phases. We think that a better model for SW Sextantis stars consists of a system with a disk wind, where the disk is fed by an overflowing accretion stream. The disk wind produces the single-peaked profiles, while the accretion stream might produce the absorption seen near phase 0.5.

We have shown that the anisotropic escape probability due to the combination of a large keplerian (disk) and poloidal (wind) velocity gradient will produce single-peaked lines from a disk. In strengthening the arguments for a disk-wind-formed line, our model also reinforces the evidence for accretion disks in nova-like variables, and in other high-luminosity objects such as quasars that show evidence for outflowing winds combined with single-peaked optical line profiles. $\square$

Received 27 March; accepted 9 July 1996.

1. Young, P., Schneider, D. P. & Shectman, S. A. *Astrophys. J.* **245**, 1035–1042 (1981).
2. Horne, K. & Marsh, T. R. *Mon. Not. R. Astron. Soc.* **218**, 761–773 (1986).
3. Young, P., Schneider, D. P. & Shectman, S. A. *Astrophys. J.* **244**, 259–268 (1981).
4. Honeycutt, R. K., Schlegel, E. M. & Kaitchuck, R. H. *Astrophys. J.* **302**, 388–402 (1986).
5. Thorstensen, J. R. et al. *Astron. J.* **102**, 272–283 (1991).
6. Dhillon, V. S., Marsh, T. R. & Jones, D. H. P. *Mon. Not. R. Astron. Soc.* **252**, 342–356 (1991).
7. Beuermann, K. et al. *Astron. Astrophys.* **256**, 442–446 (1992).
8. Dhillon, V. S., Jones, D. H. P., Marsh, T. R. & Smith, R. C. *Mon. Not. R. Astron. Soc.* **258**, 225–240 (1992).
9. Still, M. D., Dhillon, V. S. & Jones, D. H. P. *Mon. Not. R. Astron. Soc.* **273**, 863–876 (1995).
10. Marsh, T. R. & Horne, K. *Astrophys. J.* **349**, 593–607 (1990).
11. Dhillon, V. S., Marsh, T. R. & Jones, D. H. P. in *Accretion Powered Compact Binaries* (ed. Mauche, C.) 127–132 (Cambridge Univ., 1990).
12. Rutten, R. G. M. et al. *Nature* **362**, 518–520 (1993).
13. Warner, B. *Cataclysmic Variable Stars* (Cambridge Univ., 1995).
14. Williams, R. E. *Astron. J.* **97**, 1752–1758 (1989).
15. Miyoshi, M. et al. *Nature* **373**, 127–129 (1995).
16. Eracleous, M. & Halpern, J. P. *Astrophys. J. Suppl. Ser.* **90**, 1–30 (1994).
17. Kinney, A. L. in *The First Stromlo Symposium: The Physics of Active Galactic Nuclei* (eds Bicknell, B. V., Dopita, M. A. & Quinn, P. J.) 61–71 (ASP Conf. Ser. 54, San Francisco, 1994).
18. Murray, N., Chiang, J., Grossman, S. A. & Voit, G. M. *Astrophys. J.* **451**, 498–509 (1995).
19. Murray, N. & Chiang, J. *Astrophys. J.* (in the press).
20. Drew, J. E. *Mon. Not. R. Astron. Soc.* **224**, 595–632 (1987).
21. Mauche, C. W. et al. *Astrophys. J.* **424**, 347–369 (1994).
22. Shlosman, I., Vitello, P. & Mauche, C. W. *Astrophys. J.* **461**, 377–385 (1996).
23. Shlosman, I. & Vitello, P. *Astrophys. J.* **409**, 372–386 (1993).
24. van Teeseling, A. & Verbunt, F. *Astron. Astrophys.* **292**, 519–533 (1994).
25. Richman, H. R. *Astrophys. J.* **462**, 404–427 (1996).
26. Mauche, C. W., Raymond, J. C. & Matei, J. A. *Astrophys. J.* **446**, 842–851 (1995).
27. Mauche, C. W. in *Proc. IAU Colloquium 158* (eds Evans, A. & Wood, J. H.) (Dordrecht, Kluwer, 1996).
28. Horne, K. *Astron. Astrophys.* **297**, 273–284 (1995).
29. Hellier, C. & Robinson, E. L. *Astrophys. J.* **431**, L107–L110 (1994).

ACKNOWLEDGEMENTS. This work was supported by NSERC of Canada and by the Connaught Fund of the University of Toronto.

CORRESPONDENCE should be addressed to N.M. (e-mail: murray@cita.utoronto.ca).

Calorimetric measurement of the latent heat of vortex-lattice melting in untwinned $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$

A. Schilling*, R. A. Fisher*, N. E. Phillips*, U. Welp†, D. Dasgupta†, W. K. Kwok† & G. W. Crabtree†

* Lawrence Berkeley National Laboratory and Department of Chemistry, University of California, BG4 Giannini, Berkeley, California 94720-1460, USA

† Materials Science Division & Science and Technology Center for Superconductivity, Argonne National Laboratory, Argonne, Illinois 60439, USA

THE magnetic vortex lattice of copper oxide superconductors in the mixed (field-penetrated) state ‘liquefies’^{1,2} on increasing the temperature T or the external magnetic field H , giving rise to an ohmic resistivity well below the fluctuation-dominated crossover to the normal state at the upper critical field $H_{c2}(T)$. Theoretical work suggests that in clean materials this melting is a first-order phase transition³; features in the resistivity^{4,6} and magnetization^{7–10}, as well as results from muon spin rotation¹¹ and neutron-diffraction work¹², have been cited to support this hypothesis. A calorimetric measurement of a latent heat provides the most definitive proof of the occurrence of a first-order transition, but such measurements require very high sensitivity. Here we report calorimetric measurements on an untwinned single crystal of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ that have sufficient precision to clearly resolve the latent heat. The value obtained, $\sim 0.45 k_B T$ per vortex per superconducting layer (where k_B is the Boltzmann constant), is consistent with that inferred from magnetization data using the Clapeyron equation. This result is compelling evidence for a first-order transition at a well defined phase boundary $H_m(T)$.

Early experimental indications^{4,6} of the occurrence of a first-

order transition in the vortex state of untwinned $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ came from a sharply vanishing in-plane resistivity which showed hysteresis along a phase boundary $H_m(T)$ (with H parallel to the c axis of the crystal). However, the relevance of this non-equilibrium property to the nature of the phase transition was seriously questioned¹³. Discontinuities ΔM in the magnetization M of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ (refs 7, 8) and $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (refs 9, 10) provided more persuasive evidence for the occurrence of a first-order transition in the vortex state of copper oxide superconductors; however, a plausible alternative hypothesis involving effects due to an inhomogeneous (non-equilibrium) magnetic flux distribution has been proposed to explain, at least partly, the results for $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8-x}$ (ref. 14). For thermodynamic equilibrium, ΔM is related to a latent heat L and a corresponding discontinuity in entropy ΔS by the Clapeyron equation, $L = T\Delta S = -T\Delta M dH_m/dT$. Successful, direct measurements of L have not yet been reported for any copper oxide superconductor. A calorimetric investigation on a twinned $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ crystal¹⁵ gave only an upper limit for the latent heat, $L < 0.05 k_B T$ per vortex per superconducting layer. However, experimental experience shows that the presence of twin boundaries suppresses the occurrence of effects that are thought to be closely related to the first-order melting of the vortex lattice⁵. A corresponding calorimetric measurement on an untwinned material which shows both resistive and magnetization anomalies at $H_m(T)$, is therefore of considerable interest.

We used a differential thermal analysis apparatus¹⁵ to obtain calorimetric data with sufficient precision to allow both the low mass of the crystals to be investigated, and the expected small magnitude of the effect to be measured. An important feature of this technique is that it makes an unambiguous distinction between a latent heat, and thermal effects which are associated with higher-order phase transitions. We mounted a 3.3-mg, untwinned, single crystal of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ ($T_c = 92$ K) with its c axis parallel to the external magnetic field, with a reference sample (2.65 mg of copper), on cotton filaments and thermally connected them to a heat reservoir with variable temperature T_b (see Fig. 2 left inset). The crystal used was larger but otherwise almost identical, in its chemical and superconducting properties, to the sample investigated in ref. 10. Increasing H parallel to the c direction produced distinct steps ΔM in the mixed-state magnetization of the crystal (Fig. 1). We used copper–Constantan differential thermocouples to measure the differences between the temperature T_s of the superconductor (heat capacity C_s) and

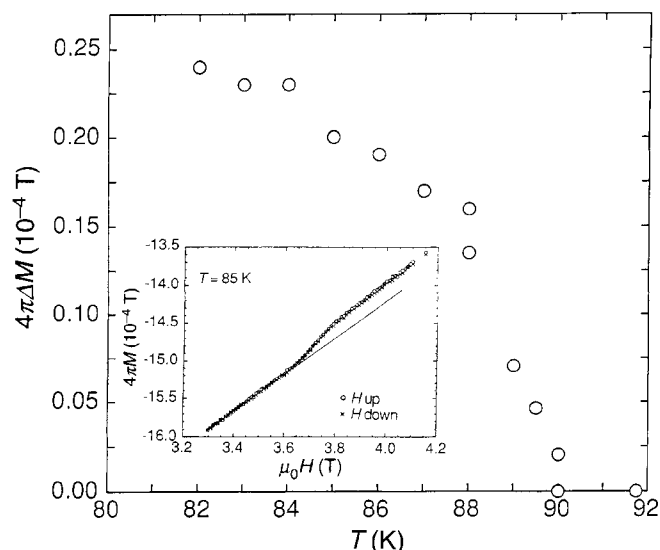


FIG. 1 Magnetization jumps $4\pi\Delta M$ versus temperature T detected on the untwinned $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ single crystal. The inset shows representative raw data to illustrate the procedure for extracting $4\pi\Delta M$ from such data.

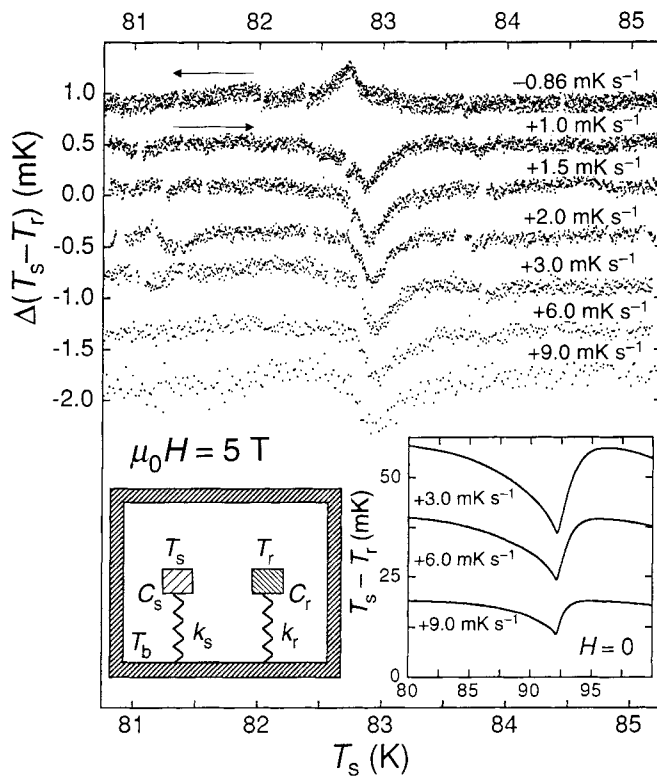


FIG. 2 Temperature difference $\Delta(T_s - T_r)$ as a function of sample temperature T_s between the untwinned $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ single crystal and a copper reference, measured in a magnetic field of $\mu_0 H = 5$ T at various heating rates. The data were vertically shifted for clarity after corrections for the smooth background differences in $C_s - C_r$ which are qualitatively similar to the 80–85 K segments displayed in the right inset. The right inset shows corresponding data taken in zero magnetic field around T_c . The left inset illustrates the experimental configuration with the heat links k_s and k_r , respectively.

the temperature T_r of the reference sample (heat capacity C_r) or T_b . The thermal time constants, determined by C_s and C_r and the thermal conductivities of the heat links, were adjusted to be similar and are typically $\tau \approx 90$ s. The contribution of the thermocouples to the total heat capacity ($\sim 25\%$) was measured in a separate experiment. All the data were corrected to account for 0.3 mg of GE 7031 varnish which was added to attach the thermocouples and fix the crystal.

As a test for the existence of a latent heat, $T_s - T_r$ was monitored for a series of different rates of change of temperature (dT_b/dt). The results were corrected for the smooth background differences between C_s and C_r , and some of them are shown in Fig. 2 as $\Delta(T_s - T_r)$. If the sample undergoes a first-order transition on heating, the increase in T_s would be interrupted and T_s would then remain constant until the heat input had supplied the latent heat. During this interval, the reference specimen would warm by $\Delta T = L/C_s$ and cause a drop in $T_s - T_r$. In a cooling experiment, $T_s - T_r$ would increase by the same amount which corresponds to the release of latent heat. Such a process would occur ideally within a time $\Delta t = \Delta T/(dT_b/dt)$, after which $\Delta(T_s - T_r)$ would relax to zero with the time constant τ , corresponding to a temperature change $T_r = \tau dT_b/dt$ (ref. 15). The expected shift in $\Delta(T_s - T_r)$ is not only independent of the heat input power, the heat capacity of the reference and the thermal constants of the heat links, but in particular the shift does not depend on the rate of change of temperature¹⁵. This is exactly what we observed. Figure 2 shows representative data obtained in a magnetic field of $\mu_0 H = 5$ T using different heating rates. In Fig. 2 the changes in $\Delta(T_s - T_r)$ at $T \approx 82.8$ K, which have almost identical amplitudes, were detected reproducibly. In contrast, an anomaly in C_s

due to a higher-order phase transition would produce shifts in $\Delta(T_s - T_r)$ with amplitudes proportional to dT_b/dt and to $C_r - C_s$ (ref. 15). These are illustrated by data measured in zero magnetic field around the second-order transition to superconductivity (Fig. 2 right inset).

Another procedure to investigate the order of the phase transition is to fix the temperature T_b and to vary the external magnetic field H . When crossing the first-order phase boundary from below, the sample must spontaneously cool by ΔT because $L > 0$. This is an effect that would be difficult to explain without invoking the occurrence of a first-order transition. Corresponding raw data are shown in the inset of Fig. 3. At $T = 83.1$ K, the sample cools at a magnetic field $\mu_0 H \approx 4.85$ T as H increases, and warms by the same amount when H is decreased. We collected similar data at various temperatures, and plotted some of the results in Fig. 3. The relative displacements of the peaks in $\Delta(T_s - T_r)$ when crossing the phase boundary in different directions, in both Figs 2 and 3, should not be taken as a measure of hysteretic effects. The shape of the features in the plots of $\Delta(T_s - T_r)$, does suggest a slight intrinsic broadening of the transition. This together with the relaxation time τ determines the details of the shape including the position of the peak.

From both types of experiment we may compute the latent heat $L = T\Delta S = \Phi_0 \Delta T C_s c / \mu_0 H k_B T$ in units of $k_B T$ per vortex per superconducting layer (here assumed to extend over the length of one crystallographic unit cell, $c = 11.7$ Å), where $\Phi_0 = 2.07 \times 10^{-15}$ Vs. The ΔT shifts, observed at constant T by varying H , must be corrected to compensate for the relaxation of $\Delta(T_s - T_r)$ while the field is slowly increased, typically at a rate $\mu_0 dH/dt = 1.67$ mT s⁻¹, across the transition (which is apparently broadened by $\mu_0 dH \approx 0.1$ T, see insets of Figs 1 and 3). The corresponding decay constant in magnetic-field units is $\mu_0 H_c = \tau \mu_0 dH/dt \approx 0.15$ T. The observed ΔT depends on H_c and amounts to $H_c / \delta H [1 - \exp(-\delta H / H_c)] \approx 75\%$ of the true value that would be observed in a perfectly adiabatic experiment (Fig. 3 inset). The results for ΔS are shown in Fig. 4 with an independent estimate calculated from the corresponding magnetization steps ΔM in Fig. 1 using the Clapeyron equation. Although the possible drop in ΔS very near T_c , suggested by the latter data, could not be

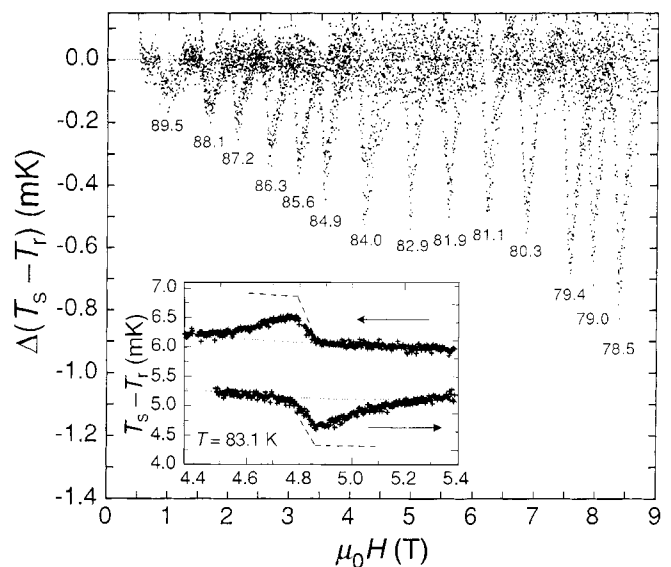


FIG. 3 Representative shifts in $\Delta(T_s - T_r)$ due to the spontaneous cooling of the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ crystal when the external magnetic field H is increased at constant temperature. Each data set is labelled by the temperature (in kelvin) at which the feature occurs. The inset shows the raw data, $T_s - T_r$, of experiments where H was increased and decreased. The dotted line indicates the background which was subtracted to obtain the net shifts $\Delta(T_s - T_r)$. The dashed lines show what might be observed under perfectly adiabatic conditions (see text).

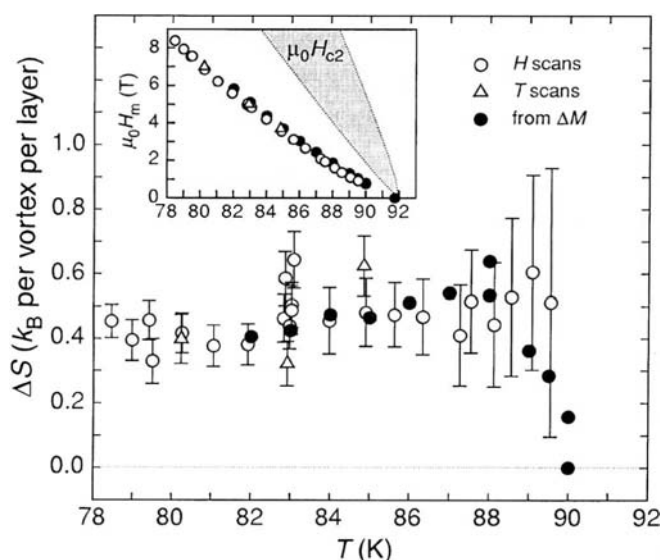


FIG. 4 Entropy change ΔS per vortex per superconducting layer which is associated with the first-order transition observed in our experiment (open symbols). The filled circles are results of an independent estimate for ΔS from magnetization shifts (ΔM) at $H_m(T)$ on the same sample. The first-order phase boundary $H_m(T)$ is shown in the inset, with a sketch of the $H_{c2}(T)$ crossover region that separates the normal and the vortex-fluid states.

resolved in the present thermal experiment, the close agreement of the two sets of data with $\Delta S \approx 0.45 k_B$ per vortex per superconducting layer ($L/\mu_0 H \approx 35 \mu J g^{-1} T^{-1}$) proves thermodynamic consistency. The location of this first-order transition in the H - T plane is shown in the inset of Fig. 4.

Our value for ΔS compares well with previous estimates ($\Delta S \approx 0.6$ – $0.8 k_B$, refs 9, 10) based on discontinuities in the magnetization at $H_m(T)$ of $YBa_2Cu_3O_{7-\delta}$. Although the magnitude of ΔS is also comparable to what has been inferred for $Bi_2Sr_2CaCu_2O_{8+x}$ (ref. 8) from magnetization data, its temperature dependence is qualitatively different. For $YBa_2Cu_3O_{7-\delta}$ we obtain a virtually constant value $\Delta S \approx 0.45 k_B$ (with a possible decrease near T_c), while in $Bi_2Sr_2CaCu_2O_{8+x}$, ΔS increases greatly as T approaches T_c , reaching values as high as $\Delta S \approx 6 k_B$ (ref. 16). This may indicate that the detailed microscopic dynamics of the first-order transition in the more anisotropic $Bi_2Sr_2CaCu_2O_{8+x}$ is fundamentally different from that in $YBa_2Cu_3O_{7-\delta}$.

Our data clearly demonstrate that a first-order phase transition takes place in the vortex state of untwinned $YBa_2Cu_3O_{7-\delta}$. The coincidence of the calorimetric result with magnetic and transport data¹⁰ strongly suggests that this transition is due to the first-order melting of the vortex lattice. This transition, with its obvious change in translational symmetry, is probably the only true phase transition of the electronic subsystem of copper oxide superconductors near T_c and for magnetic fields H exceeding the lower critical field $H_{c1}(T)$. The upper critical field $H_{c2}(T)$, which separates the superconducting and normal states in other classes of type II superconductors with a second-order phase transition, is replaced in the copper oxides by a fluctuation-dominated continuous crossover.² □

Received 6 June; accepted 23 July 1996.

1. Nelson, D. R. *Phys. Rev. Lett.* **60**, 1973–1976 (1988).
2. Fisher, D. S., Fisher, M. P. A. & Huse, D. A. *Phys. Rev. B* **43**, 130–159 (1991).
3. Hetzel, R. E., Sudbø, A. & Huse, D. A. *Phys. Rev. Lett.* **69**, 518–521 (1992).
4. Safar, H. et al. *Phys. Rev. Lett.* **69**, 824–827 (1992).
5. Kwok, W. K. et al. *Phys. Rev. Lett.* **69**, 3370–3373 (1992).
6. Kwok, W. K. et al. *Phys. Rev. Lett.* **72**, 1092–1095 (1994).
7. Pastoriza, H., Goffman, M. F., Amibère, A. & de la Cruz, F. *Phys. Rev. Lett.* **72**, 2951–2954 (1994).
8. Zeldov, E. et al. *Nature* **375**, 373–376 (1995).
9. Liang, R., Bonn, D. A. & Hardy, W. N. *Phys. Rev. Lett.* **76**, 835–838 (1996).

10. Welp, U., Fendrich, J. A., Kwok, W. K., Crabtree, G. W. & Veal, B. W. *Phys. Rev. Lett.* **76**, 4809–4812 (1996).
11. Lee, S. L. et al. *Phys. Rev. Lett.* **71**, 3862–3865 (1993).
12. Cubitt, R. et al. *Nature* **365**, 407–411 (1993).
13. Jiang, W., Yeh, N. C., Reed, D. S., Kriplani, U. & Holtzberg, F. *Phys. Rev. Lett.* **74**, 1438–1441 (1995).
14. Farrell, D. E. et al. *Phys. Rev. B* **53**, 11807–11816 (1996).
15. Schilling, A. & Jeandupeux, O. *Phys. Rev. B* **52**, 9714–9723 (1995).
16. Morozov, N., Zeldov, E., Majer, D. & Konczykowski, M. *Phys. Rev. B* **54**, 3784 (1996).

ACKNOWLEDGEMENTS. A.S. would like to thank to the Schweizerische Nationalfonds zur Förderung der Wissenschaftlichen Forschung for current and previous support. The work at Berkeley and at Argonne was supported by the US Department of Energy. D.D. was supported by the US National Science Foundation Science and Technology Center for Superconductivity. We also thank D. Wright for useful assistance.

CORRESPONDENCE should be addressed to A.S. (e-mail: schilling@solid.phys.ethz.ch).

Localized excitations in a vertically vibrated granular layer

Paul B. Umbanhowar*, Francisco Melo† & Harry L. Swinney*

* Center for Nonlinear Dynamics and Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA
† Departamento de Física, Universidad de Santiago, Avenida Ecuador 3493, Casilla 307 Correo 2, Santiago, Chile

THE formation of two-dimensional patterns in biological, chemical and physical systems is often described by the nonlinear interaction of plane waves¹. An alternative approach views patterns as ensembles of interacting localized objects, analogous to the assembly of crystals from atoms. For macroscopic pattern-forming systems, one objection to the latter approach is that no ‘atoms’ exist; however spatially localized excitations can play an analogous role. One-dimensional localized states are observed in many systems—for example, solitary waves in water^{2–4} and optical fibres⁵—and can organize into simple patterns^{6,7}. But few examples of two-dimensional localized states are known, and these tend to be unstable and/or do not show simple pattern-forming interactions^{8–11}. Here we report the observation of stable, two-dimensional localized excitations in a vibrating layer of sand. These excitations, which we term ‘oscillons’, have a propensity to assemble into ‘molecular’ and ‘crystalline’ structures. Our experimental results, together with the observation of similar localized excitations in model differential equations^{12–14}, indicate a crucial, cooperative role for hysteresis and dissipation in the formation of oscillons, and suggest that similar behaviour may occur in continuous media.

The oscillons in our study are found in granular materials—highly dissipative media composed of macroscopic grains that interact by means of contact forces. Despite substantial interest, granular materials remain poorly understood¹⁵. To maintain motion in these materials an energy source is required. Experiments in which the energy is supplied by a vertically vibrated plate show heap formation and convection^{16,17}, size segregation¹⁸, bubbling¹⁹ and subharmonic standing waves^{20–22}. All but the last of these phenomena are caused by the surrounding gas and/or side-wall friction²³. Granular standing waves, however, arise solely from cooperative behaviour resulting from collisions between particles. Granular waves can be stripes, squares or hexagons, depending on the vibration frequency and amplitude²².

Oscillons form under the same conditions that are required for granular waves. In our experiment, a granular layer in the bottom of an evacuated, upright, cylindrical container²² is vertically driven with a displacement (as a function of time t) $z = A \sin(2\pi ft)$, where A is the amplitude of the displacement. Our control parameters are the drive frequency f and the dimensionless acceleration

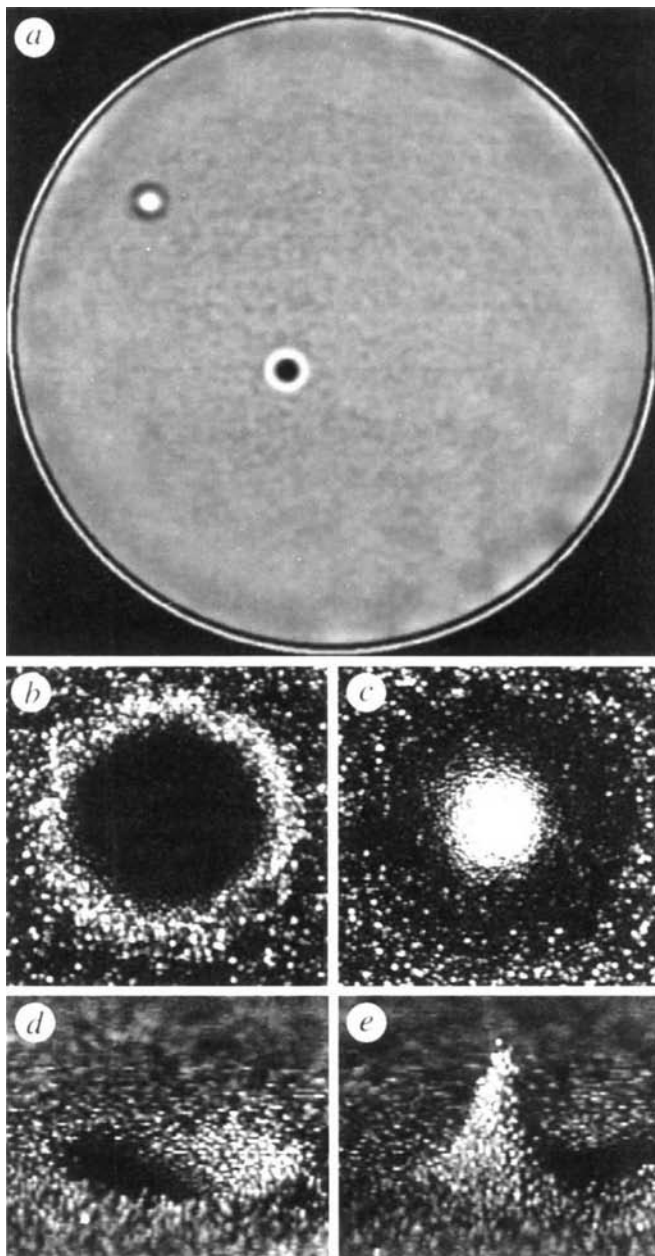


FIG. 1 *a*, Two oscillons with opposite phase. *b* and *c*, A single oscillon viewed from above, at times differing by $1/f$. *d* and *e*, Corresponding side views ($f = 26$ Hz, $\Gamma = 2.45$, layer depth of 17 particles). The individual particles (0.15–0.18-mm-diameter bronze spheres) are in close contact, as can be seen in *b–e*, which show $7\text{ mm} \times 7\text{ mm}$ regions of the 127-mm-diameter container. Images are illuminated from the side, for $1/100$ of a cycle, with light strobed in phase with the drive signal so that regions with large amplitude are light, and regions with small amplitude are dark. The container is evacuated to 0.1 torr, a value at which the effects of the remaining gas are negligible.

amplitude, $\Gamma = 4\pi^2 A f^2 g^{-1}$ (where g is the acceleration due to gravity).

An oscillon is a small, circularly symmetric excitation which oscillates at frequency $f/2$: during one cycle of the container, it is a peak; on the next cycle it is a crater (Fig. 1). Because oscillons are subharmonic, peaks and craters can coexist. Examining the layer only on alternate cycles shows that craters continue to be craters and peaks continue to be peaks. The diameter of oscillons depends on Γ but is typically 30 particles (4% of the container diameter). Oscillons form with equal probability at all locations within the plane of the layer and are long lived; they often persist

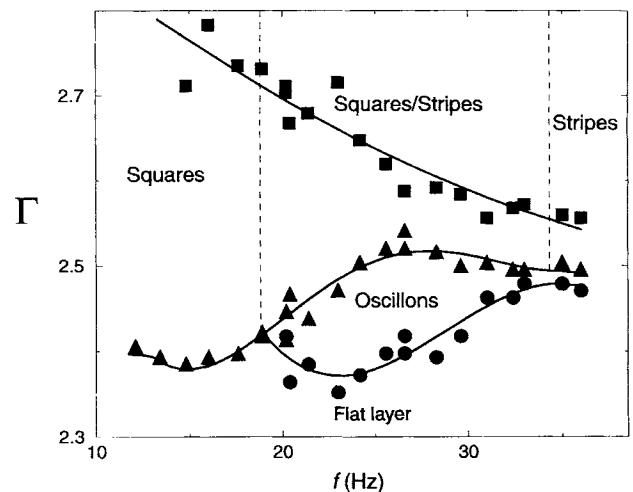


FIG. 2 Diagram showing the stability regions for different states, as a function of f and Γ , for increasing Γ (squares) and decreasing Γ (triangles and circles). The transitions from the flat layer to squares and stripes are hysteretic, but the hysteresis is much smaller for stripes. Oscillons are observed for layers greater than 13 particles deep in a range of f which increases with increasing depth. For thinner layers, the phase diagram is similar but without the oscillon region.

for more than 5×10^5 container oscillations. In essence, oscillons do not propagate, but they do drift about slowly, and randomly, taking about 10^3 cycles to move a distance equal to their diameter.

Figure 1*b* and *d* shows a single oscillon which, at this point in the cycle, is a circular depression (crater) surrounded by a raised ring of displaced material in an otherwise flat layer. One cycle later, as shown in Fig. 1*c* and *e*, the oscillon is a peak extending to a height of about four times the layer depth. In both phases, the oscillon is formed by material displaced from its immediate vicinity.

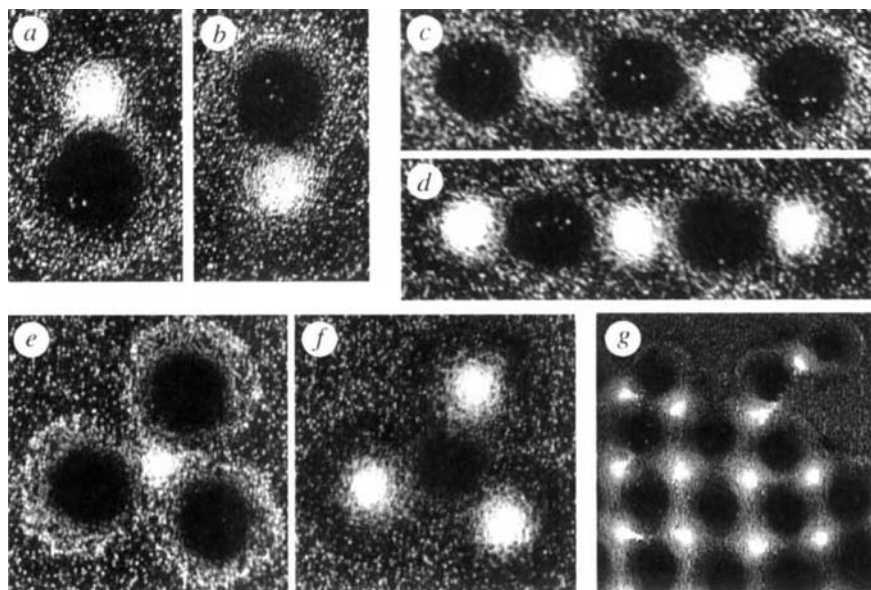
For the parameter range where oscillons are found, the granular layer undergoes a free flight which begins when the container acceleration becomes less than $-g$, followed by an impulsive acceleration when the layer collides with the base of the container. Oscillons change from craters to peaks (or peaks to craters) during the collision portion of the cycle. When an oscillon in the crater phase collides with the container, grains are accelerated towards the centre and form a peak during the subsequent free flight. At the next impact, particles flow outwards and reform the crater.

The free-flight time and collision velocity of the layer are determined by Γ (ref. 22). For $\Gamma < 1$ the layer remains on the plate; for $\Gamma > 1$ the layer undergoes free flight but remains flat up to the point where wave patterns appear at $\Gamma \approx 2.5$, as Fig. 2 shows. Square patterns form for $f < 18$ Hz and show hysteresis—the value of Γ at which squares appear from the flat layer, for increasing acceleration, is about 20% larger than the value of Γ at which they disappear for decreasing acceleration. For $f > 35$ Hz, stripe patterns form with a hysteresis of about 5% at 35 Hz and decreasing hysteresis for increasing f . For intermediate values of f , both patterns are present and interact in a complicated time-dependent manner.

Oscillons are found in a range of Γ below the lower stability boundary of the standing-wave patterns for frequencies in the transition region between squares and stripes. Oscillons do not form spontaneously from the flat layer. They are produced by locally perturbing the layer or by decreasing Γ from a patterned state. In the latter case, oscillons frequently form at pattern defects. The number of oscillons for a given f and Γ is not fixed; quickly increasing Γ from the flat layer state into the square and stripe state, and then reducing Γ into the oscillon regime can produce from 0 to about 50 oscillons.

As Γ is decreased from the upper to the lower stability boundary for oscillons, the oscillon diameter decreases. At $f = 26$ Hz, it is about 40% smaller at the lower stability boundary than it is at the

FIG. 3 Oscillon 'molecules' and 'crystal': a, b, dipole; c, d, polymeric chain; e, f, triangular tetramer; g, square ionic lattice. The pairs of pictures in a–f are separated in time by $1/f$. In all pictures the separation between adjacent oscillons of opposite phase is approximately 4 mm. In g, a small chain separates from the edge of the square lattice.



upper stability boundary. When Γ is decreased below the lower oscillon stability boundary, the diameter decreases to zero over 5–10 cycles, and the excitation vanishes. When Γ is increased above the upper stability boundary for oscillons, two different processes occur, depending on the magnitude of f . For $f < 25$ Hz oscillons nucleate oscillons of opposite phase at their outer diameters. For $f > 30$ Hz oscillons elongate along one axis and nucleate wavefronts along the other. For intermediate f , both processes are observed. In all cases, the final state is composed of patterns which continuously change back and forth between squares and stripes.

Oscillons of like phase show a short-range, repulsive interaction, while oscillons of opposite phase attract and bind. When a peak and a crater collide, they attract, overlap and form a stable dipole structure, as shown in Fig. 3a, b. When joined, oscillons retain their shape and relative phase, despite being only about one radius apart. The interaction between oscillons is short-ranged—oscillons with centres separated by more than about 1.4 oscillon diameters behave independently.

In addition to dipoles, other stable combinations of peaks and craters exist for which the constituent oscillons have coordination numbers greater than one. Figure 3c, d shows a chain in which the internal oscillons have coordination number two. The chains are straight, which indicates that bound oscillons influence each other even when they are not in direct contact. This interaction is possibly mediated by the intervening links in the chain. Triangular tetramer structures also form and in these, the central oscillon has coordination number three (Fig. 3e, f).

Isolated compound structures with coordination number four or higher are not observed. Square patterns, however, show interesting behaviour when Γ is decreased slightly below the lower stability boundary for squares. In this case, the flat layer invades the square pattern over the course of hundreds of cycles; as it invades, small chains and individual oscillons separate from

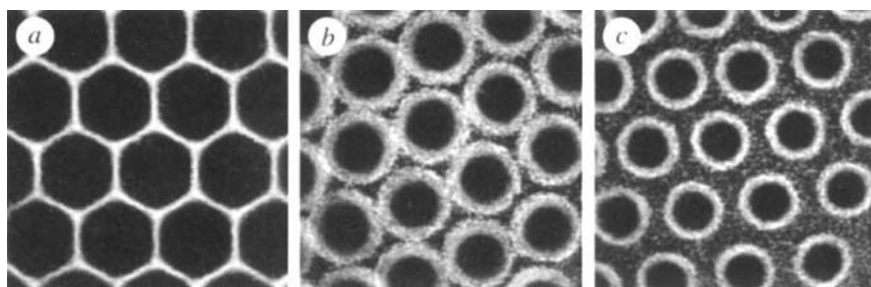
the lattice and then disappear (Fig. 3g). This process suggests that in this region, stable, planar, square patterns are composed of peaks and craters in an arrangement similar to an ionic crystal.

Additional evidence, showing that planar patterns can in some cases to be considered as collections of oscillons is provided by experiments with an added subharmonic forcing term at frequency $f/2$. The time translation symmetry $t \rightarrow t + f^{-1}$ is broken and planar hexagonal patterns form where squares and stripes appeared previously (Fig. 4). When Γ is decreased into the oscillon region, the hexagonal pattern becomes a hexagonal crystal of oscillons with identical phase. These oscillons do not bind together as they are weakly repelling, as is the case for single frequency oscillons with the same phase. Increasing Γ from this state makes the oscillons expand until the cellular honeycomb pattern is recovered.

There is no equation of motion for granular media that could produce a quantitative explanation of the formation and mutual interaction of oscillons. However, a qualitative explanation based on the roles of hysteresis and dissipation in localized structure formation¹² provides much insight. For oscillons to exist, the primary bifurcation must be hysteretic so that waves and the flat layer can coexist at the same parameter values. In our experiments, hysteresis is large for low f and small for high f . Oscillons, however, appear only in a range of f where the hysteresis is decreasing with increasing frequency (Fig. 2). We suggest that dissipation, which keeps structures localized, is too small to stabilize oscillons for low f .

To investigate the possible role of dissipation, we conducted experiments using particles with different diameters D . The frequency at the centre of the range in which oscillons occur, was found to be proportional to $D^{-1/2}$. This scaling can be easily understood: for large ratios of the relative kinetic energy of a particle to the potential energy needed to raise the particle by one

FIG. 4 Hexagonal patterns with two-frequency forcing, $z' = A \sin(2\pi f_1 t) + B \sin(2\pi f_2 t + \pi/2)$ with $B/A = 0.04$, $f_1 = 26$ Hz, and $f_2 = 13$ Hz. a, Above the planar pattern threshold ($\Gamma = 2.70$), b, at the threshold ($\Gamma = 2.50$), and c, below the threshold ($\Gamma = 2.45$). Each frame is 40 mm $\times$ 40 mm. In b, many cells are still connected, although regions of flat layer have appeared between them. In c, the pattern is a crystal of weakly repelling oscillons.



diameter, that is v^2/gD (where v is the vertical velocity relative to a neighbouring particle), the horizontal mobility is high and the dissipation is small²⁴. We assume that Γ is constant and that v is proportional to the maximum container velocity (so that $v \propto f^{-1}$); then at low f , $v^2/gD \gg 1$ and the dissipation is small, while at large f , $v^2/gD \ll 1$ and the dissipation is large. The observed scaling of the centre frequency for oscillon formation with D is consistent with a constant value of v^2/gD or, equivalently, a constant dissipation. Thus, oscillons form only in the narrow range of f where both dissipation and hysteresis are large. Support for the dependence of dissipation on f also comes from the observation that in vertically vibrated fluids (the Faraday system), square patterns appear at low viscosity while stripes occur at high viscosity²⁵; this is

the same transition observed in our experiment as f is increased through the oscillon region.

Numerical simulations of model systems with hysteresis and dissipation also show localized two-dimensional structures^{12–14}. Pairs of localized excitations can form bound states similar to our oscillon dipoles: the centres of adjacent localized excitations are separated by slightly more than one radius; individual solutions maintain their identity after binding; and paired structures are stable¹³. Our experiments and these simulations indicate the crucial, cooperative role of hysteresis and dissipation in the formation of two-dimensional localized structures, and suggest that similar behaviour may also occur in other physical systems with these properties. □

Received 5 March; accepted 15 July 1996.

1. Cross, M. C. & Hohenberg, P. C. *Rev. Mod. Phys.* **65**, 851–1112 (1993).
2. Russell, J. S. *Report on Waves* 311–390 (Murray, London, 1844).
3. Wu, J., Keolian, R. & Rudnick, I. *Phys. Rev. Lett.* **52**, 1421–1424 (1984).
4. Osborne, A. R. & Burch, T. L. *Science* **208**, 451–460 (1980).
5. Mollenauer, L. F., Stolen, R. H. & Gordon, J. P. *Phys. Rev. Lett.* **45**, 1095–1098 (1980).
6. Melo, F. & Douady, S. *Phys. Rev. Lett.* **71**, 3283–3286 (1993).
7. Balmforth, N. J. *Annu. Rev. Fluid Mech.* **27**, 337–373 (1995).
8. Dennin, M., Ahlers, G. & Cannell, D. S. *Science* **272**, 388–390 (1996).
9. Joets, A. & Ribotta, R. *Phys. Rev. Lett.* **21**, 2164–2167 (1988).
10. Lerman, K., Bodenschatz, E., Cannell, D. S. & Ahlers, G. *Phys. Rev. Lett.* **70**, 3572–3575 (1993).
11. Lioubashevski, O., Arbell, H. & Fineberg, J. *Phys. Rev. Lett.* **76**, 3959–3962 (1996).
12. Thual, O. & Fauve, S. *J. Phys., Paris* **49**, 1829–1833 (1988).
13. Deissler, R. J. & Brand, H. R. *Phys. Rev. A* **44**, R3411–R3414 (1991).
14. Aranson, I. S., Gorshkov, K. A., Lomov, A. S. & Rabinovich, M. I. *Physica D* **43**, 435–453 (1990).

15. Jaeger, H. M., Nagel, S. R. & Behringer, R. P. *Physics Today* **49**(4), 32–38 (1996).
16. Evesque, P. & Rajchenbach, J. *Phys. Rev. Lett.* **50**, 44–47 (1989).
17. Laroche, C., Douady, S. & Fauve, S. *J. Phys., Paris* **50**, 699–702 (1989).
18. Knight, J. B., Jaeger, H. M. & Nagel, S. R. *Phys. Rev. Lett.* **70**, 3728–3730 (1993).
19. Pak, H. K. & Behringer, R. P. *Nature* **371**, 231–233 (1994).
20. Fauve, S., Douady, S. & Laroche, C. *J. Phys. Colloque, Paris* **50** C3, 187–191 (1989).
21. Melo, F., Umbanhowar, P. & Swinney, H. L. *Phys. Rev. Lett.* **72**, 172–175 (1994).
22. Melo, F., Umbanhowar, P. & Swinney, H. L. *Phys. Rev. Lett.* **75**, 3838–3841 (1995).
23. Pak, H. K., Van Doorn, E. & Behringer, R. P. *Phys. Rev. Lett.* **74**, 4643–4646 (1995).
24. Reynolds, O. *Phil. Mag.* **20**, 469–481 (1885).
25. Edwards, W. S. & Fauve, S. *J. Fluid Mech.* **278**, 123–148 (1994).

ACKNOWLEDGEMENTS. We thank N. Balmforth and E. Weeks for helpful discussions. This research is supported by the US Department of Energy Office of Basic Energy Sciences, the Texas Advanced Research Program, the US National Science Foundation Division of International Programs and Fondecyt (Chile).

CORRESPONDENCE should be addressed to P.B.U. (e-mail: pbu@chaos.ph.utexas.edu).

A silicoboron carbonitride ceramic stable to 2,000 °C

Ralf Riedel*, Andreas Kienzle††, Wolfgang Dressler*, Lutz Ruwisch*, Joachim Bill† & Fritz Aldinger†

* Fachgebiet Disperse Feststoffe, Fachbereich Materialwissenschaft, Technische Hochschule Darmstadt, Petersenstrasse 23A, D-64287 Darmstadt, Germany

† Max-Planck-Institut für Metallforschung, Institut für Werkstoffwissenschaft, Pulvermetallurgisches Laboratorium, Heisenbergstrasse 5, D-70569 Stuttgart, Germany

CERAMICS based on silicon nitride and carbide are strong and stable at high temperatures, and are therefore under investigation for the fabrication of motor and turbine parts^{1–3}. But silicon nitride decomposes at about 1,400 °C in vacuum and 1,775 °C in 0.1 MPa nitrogen^{4,5}, limiting the high-temperature range of its technological uses. Here we describe a boron-containing silicon nitride/carbide ceramic that does not degrade at temperatures up to 2,000 °C even in nitrogen-free environments. We synthesize the material in a polymer-to-ceramic transformation⁶ from a single polymeric polyborosilazane precursor. On heating at 1,000 °C in argon we obtain a ceramic with the composition $\text{Si}_{3.0}\text{B}_{1.0}\text{C}_{4.5}\text{N}_{2.0}$. The ceramic begins to convert to a polycrystalline composite of silicon nitride and carbide (with some non-crystalline boron nitride) at 1,700 °C, a process that is completed (without substantial change in elemental composition) at 2,000 °C.

The single-source ceramic precursor, tris(dichloromethylsilyl) ethylborane (**1**; Fig. 1a), is synthesized by the dropwise addition of 0.5 mol dimethylsulphide borane, $(\text{CH}_3)_2\text{S} \cdot \text{BH}_3$ (in 250 ml toluene solution), to 1.5 mol dichloromethylvinylsilane dissolved in 300 ml toluene at 0 °C (Fig. 1a). After stirring of the reaction mixture for 36 h at room temperature and subsequent evaporation of the solvent, a viscous liquid, compound **1**, is isolated in 94% yield. According to ¹H-, ¹³C- and ²⁹Si-NMR investigations, the reaction product contains a mixture of configurational isomers

due to the formation of two chiral α -C atoms. Compound **1** is obtained by hydroboration of the silane at the vinylic α - or β -C atoms in the ratio 2:1 (Fig. 1a). This type of hydroboration process corresponds to the reaction of boranes with alkenylsilanes^{7,8}. Further Fourier transform infrared (FTIR) and NMR spectroscopic studies of **1** support a trigonal planar coordination for boron connected to three carbon atoms (¹¹B-NMR, $\delta = +84$ p.p.m.; infrared, $\nu = 1,175 \text{ cm}^{-1}$ (B–C vibration of BC_3 -group)): all three B–H bonds of the BH_3 adduct have reacted. The organosilylborane **1** is subsequently polycondensed at 0 °C in tetrahydrofuran (THF) by ammonolysis (Fig. 1a). After separation from the by-product, ammonium chloride, and distillation of the solvent at room temperature, the polyorganoborosilazane **2** (Fig. 1a) is isolated in 80% yield as a white solid powder soluble in THF and toluene. The proposed polymer structure of **2** derived from ¹¹B- and ²⁹Si-NMR studies can be considered as a network containing cyclic oligosilazanes two-dimensionally cross-linked via Si–C–C–B–C- and Si–C–B–C-bridges (Fig. 1b).

The polymer-to-ceramic transformation of the novel polyorganoborosilazane **2** was studied by thermal gravimetric analysis (TGA) between room temperature and 1,000 °C under Ar and with a heating rate of 5 °C min^{–1}. The thermally induced decomposition of **2** begins at 220 °C and is completed at 800 °C. Ceramic material (found by X-ray diffraction to be amorphous) in the quaternary system Si–B–C–N is obtained as the reaction product with 62 wt% yield after thermolysis at 1,000 °C. The solid phase density, determined by pycnometry, is 2.5 g cm^{–3} and the empirical formula is calculated from elemental analysis (Table 1) as $\text{Si}_{3.0}\text{B}_{1.0}\text{C}_{4.5}\text{N}_{2.0}$. The molar ratio B : Si = 1 : 3 of **2** remains constant during the polymer-to-ceramic transformation (1,000 °C in Ar).

The thermal stability of nitride-based ceramics is strongly dependent on the nitrogen partial pressure. The decomposition reaction of the binary compound Si_3N_4 according to equation (1)



has an equilibrium vapour pressure of 0.1 MPa (1 bar) at 1,775 °C. The volatility of silicon nitride ceramics limits their use to about 1,500 °C (refs 4, 5). Polyorganosilazane-derived amorphous silicon carbonitrides with the general composition $\text{Si}_{3-x}\text{C}_{x-y}\text{N}_y$, reveal substantial mass loss at $T \geq 1,400$ °C. The decomposition is related to the solid-state reaction of carbon with Si_3N_4 formed in

† Present address: Institut für Fertigungstechnologie für keramische Bauteile, Universität Stuttgart, Breitscheidstrasse 8, D-70174 Stuttgart, Germany.

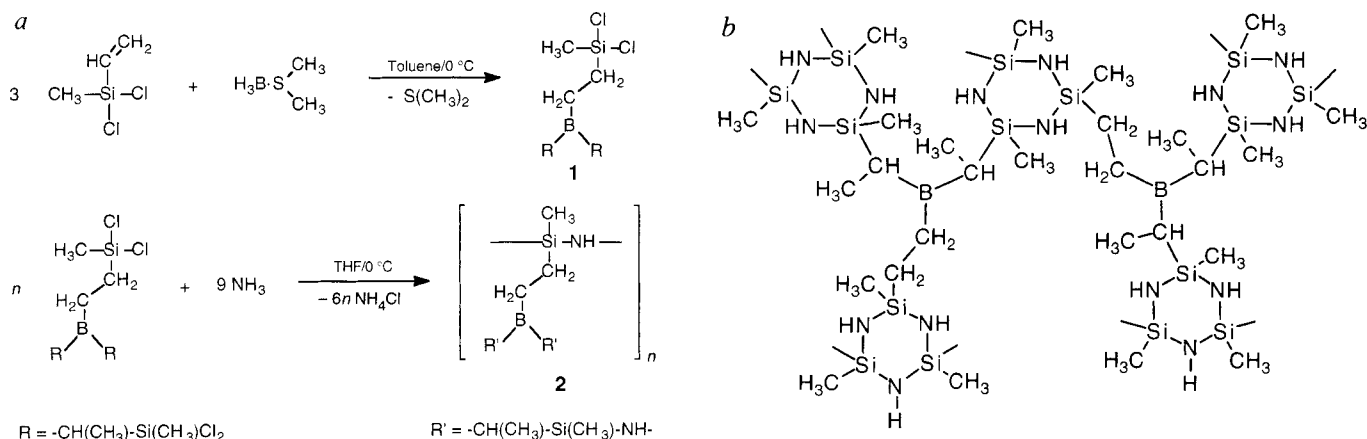


FIG. 1 **a**, Reaction path representing the synthesis of polyorganoborosilazane **2**. Polymer **2** was characterized as follows. Gel permeation chromatography with polystyrol as standard exhibits an average molecular weight $M_w = 1,099$ ($g\ mol^{-1}$) and a number-average molecular weight $M_n = 861$ ($g\ mol^{-1}$) with a polydispersity of 1.28. Elemental bulk analysis of **2** is in good accordance with the composition $Si_3B_1C_4.3N_{2.0}$ (molar weight per monomeric unit: 269). Analysis: wt% (calc.) C = 40.5 (40.2); H = 8.9

(8.9); N = 12.0 (15.6). **b**, Section of the molecular structure of polyorganoborosilazane **2** as analysed by ^{11}B - and ^{29}Si -NMR. Accordingly, the polymer structure of **2** contains cyclic methylsilazane units, $[-RSi(CH_3)_2-NH]_n$ with $n = 3-5$, cross-linked via the substituent $R = [-CH(CH_3)_2B-CH_2-CH_2-]$ attached to three silicon atoms. ^{11}B -NMR (internal standard: B_2O_3): $\delta = +75$ p.p.m.; ^{29}Si -NMR (internal standard: $(CH_3)_4Si$): resonance signals between $\delta = -10$ and $+0.5$ p.p.m.

the course of phase-partitioning or crystallization of $Si_{3+x}C_{x+y}N_4$ (refs 9, 10).

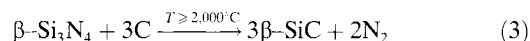
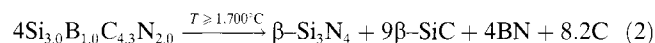
Enhanced high-temperature stability of organoborosilicon-polymer-derived ceramic fibres in comparison to conventional SiC fibres was first reported in 1986¹¹. Baldus *et al.* and Löffelholz and Jansen described the synthesis of X-ray-amorphous $Si_3B_3N_7$ (refs 12–16) and $SiBCN_3$ (ref. 15) with thermal stabilities up to 1,700 °C (0.1 MPa He) and 1,800 °C (0.1 MPa Ar), respectively. Funayama *et al.* found that polyborosilazane-derived Si–B–N–O ceramics are characterized by a higher thermal stability (1,700 °C in 0.1 MPa N_2) with respect to Si_3N_4 synthesized from perhydro-poly-silazane¹⁷.

We found that $Si_{3.0}B_{1.0}C_{4.3}N_{2.0}$ synthesized from **2** has a significantly increased thermal stability with respect to pure silicon nitride and other silicon-nitride-based materials. This phenomenon was studied by elemental bulk analysis and by analysis of the relative mass loss of the silicoboron carbonitride versus temperature compared to the behaviour of pure α - Si_3N_4 and polyhydrido-methylsilazane-derived silicon carbonitride, $Si_{1.7}C_{1.0}N_{1.6}$ (Fig. 2 and Table 1). $Si_{3.0}B_{1.0}C_{4.3}N_{2.0}$ withstands thermal degradation up to $\sim 2,000$ °C in 0.1 MPa He atmosphere. Moreover, the analysed elemental composition of the silicoboron carbonitride annealed for 2 h at 2,000 °C corresponds almost to that of the as-synthesized ceramic (1,000 °C), indicating the ultra-high thermal stability of the material. Heat-treatment at $T > 2,000$ °C results in a distinct mass loss and compositional change (Table 1). The decomposition rate of the ceramic increases significantly at $T \geq 2,100$ °C (Fig. 2). The continuous mass loss, of the order of 6 wt% between 1,400 and 2,000 °C, is caused by the evaporation of oxygen-containing species like B_2O_3 , SiO and/or CO. This result correlates with the decrease of the oxygen contamination from 3.8 wt% in the as-synthesized product to 0.3–0.5 wt% in the samples annealed at 2,000 and 2,100 °C, respectively.

The polyborosilazane-derived $Si_{3.0}B_{1.0}C_{4.3}N_{2.0}$ resists crystallization below 1,700 °C (Fig. 3). Annealing of the synthesized ceramic at 1,700 °C for 50 h in Ar results in the formation of β - Si_3N_4 and β -SiC polycrystals, while at 2,100 °C β -SiC is found as the only crystalline phase. The heat-treated samples contain 5.6–7.6 wt% boron (Table 1). Although crystalline boron-containing phases and free carbon (graphite) could not be identified in the X-ray diffractograms, amorphous or turbostratic boron nitride was clearly identified in the samples annealed at 1,700 and 2,000 °C by the characteristic BN absorption band located at $1,380\ cm^{-1}$ in the FTIR spectra.

Taking into account the IR spectroscopic and X-ray diffraction

data, the crystallization and subsequent decomposition of the metastable quaternary phase $Si_{3.0}B_{1.0}C_{4.3}N_{2.0}$ observed at $T \geq 1,700$ °C and $T \geq 2,100$ °C, respectively, can be represented by phase partitioning or crystallization (equation (2)) and decomposition (equations (3)) and these are summarized by equation (4):



Evaluation of the elemental analysis data (Table 1) gives a polycrystalline solid comprised of 2.7 β -SiC/1 BN/0.6 C, indicating a loss of carbon and silicon in addition to the formation of nitrogen gas during the solid-state decomposition at $T > 2,000$ °C. These findings clearly show that the transient solid-state reaction of carbon with silicon nitride to give silicon carbide and gaseous nitrogen

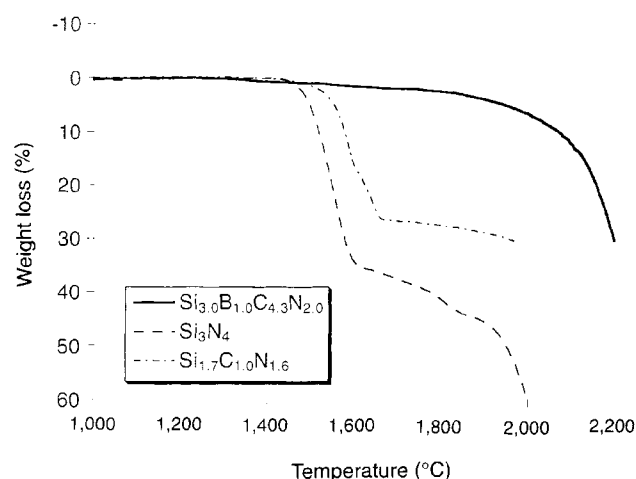


FIG. 2 Thermal gravimetric analysis (equipment; STA 429, Netzsch Gerätebau, Germany) of powdered polycrystalline α - Si_3N_4 (UBE SN-E10), silicon carbonitride ($Si_{1.7}C_{1.0}N_{1.6}$), derived from polysilazane (NCP 200, Nichimen Corp., Japan) and silicoboron carbonitride ($Si_{3.0}B_{1.0}C_{4.3}N_{2.0}$) derived from polyborosilazane **2**. Weight loss versus temperature is shown, and the data were obtained by heating under stationary 0.1 MPa He in graphite crucibles; heating rate, $5\ K\ min^{-1}$.

TABLE 1 Analysis of heat-treated silicoboron carbonitride

Annealing temp. (°C) holding time (h)	Mass loss* (wt%)	Composition (wt%)					Total	Empirical formula†
1,000/1	0	45.6	6.0	28.6	15.5	3.8	99.5	Si _{3.0} B _{1.0} C _{4.3} N _{2.0}
2,000/2	6	45.0	5.6	30.3	16.3	0.3	97.6	Si _{3.2} B _{1.0} C _{5.0} N _{2.3}
2,100/0.5	39	52.8	7.6	27.6	11.0	0.5	99.5	Si _{2.7} B _{1.0} C _{3.3} N _{1.1}

Chemical analysis of Si_{3.0}B_{1.0}C_{4.3}N_{2.0}, heat-treated at various temperatures under 0.1 MPa Ar. Silicon and boron were determined by optical emission spectrometry with inductively coupled plasma excitation (OES-ICP, Perkin Elmer ICP 5500, standard deviation <1% with Cu as internal standard); carbon, nitrogen and oxygen were analysed by carrier gas heat extraction (CGHE, LECO C,S-analyzer 244 and N, O-determinator TC 436, standard deviation <1% for C and <2% for N and O).

* Mass loss of Si_{3.0}B_{1.0}C_{4.3}N_{2.0} after isothermal annealing at 2,000 and 2,100 °C.

† Calculated without considering the oxygen contamination.

(equation (3)) is suppressed in the silicoboron carbonitride up to 2,000 °C. In contrast, boron-free silicon carbonitride^{9,10} decomposes rapidly at temperatures above 1,400 °C (Fig. 2).

The thermal stability of the silicoboron carbonitride is related to the synthesis procedure. Co-ammonolysis of **1** with (CH₃)₂SiCl₂ or reaction of **2** with BH₃·S(CH₃)₂ and subsequent thermolysis of the corresponding polyborosilazanes obtained resulted in silicoboron carbonitrides¹⁸ decomposing at $T > 1,400$ °C. This result implies that the use of single-source precursors is an essential requirement for the synthesis of Si–B–C–N materials with ultra-high thermal stability.

To have practical potential, the new material must have long-term stability and corrosion resistance at ultra-high temperatures. Preliminary experiments revealed that isothermal annealing of Si_{3.0}B_{1.0}C_{4.3}N_{2.0} at 1,600 and 1,700 °C in 0.1 MPa Ar, He or N₂ for 50 h gave no mass change. Moreover, Si_{1.7}C_{1.0}N_{1.6} and SiBN₃C were found (by analysis) to have high oxidation resistance and the lowest oxidation rate between 1,200 and 1,600 °C in comparison to chemical-vapour-deposited Si₃N₄ and SiC (refs 9, 15). These results support the compositional stability of the novel Si–B–C–N ceramics even under long-term and corrosive ultra-high temperature applications.

The novel polyborosilazane **2** can be processed to give ceramic fibres and bulk materials. Fibres are produced by dry spinning of the polymer in the THF solutions. Fibre formation is favoured by the two-dimensional structure of polymer **2**. Subsequent pyrolysis of the polymer fibres is conducted in Ar up to 1,100 °C, with a heating rate of 2 K min^{−1} and a 3-h hold at the final temperature, and yields the corresponding silicoboron carbonitride fibres. Because they are in the amorphous state, the fibres have smooth surfaces—an important requirement for reinforcement by pull-out in ceramic fibre composites¹⁹. Analogous to the procedure described in ref. 10, thermally cross-linked polyborosilazane powder can be compacted and pyrolysed to form crack-free silicoboron carbonitride bulk ceramics with ~20% open porosity as determined by Hg-pressure porosimetry.

On the one hand, the high resistance towards crystallization and thermal degradation of Si_{3.0}B_{1.0}C_{4.3}N_{2.0} can be utilized for the production of ceramic fibres, coatings and composites which are chemically and mechanically stable at ultra-high temperatures. As the new silicoboron carbonitride does not tend to crystallize up to 1,600–1,700 °C, it is expected that the initial room-temperature

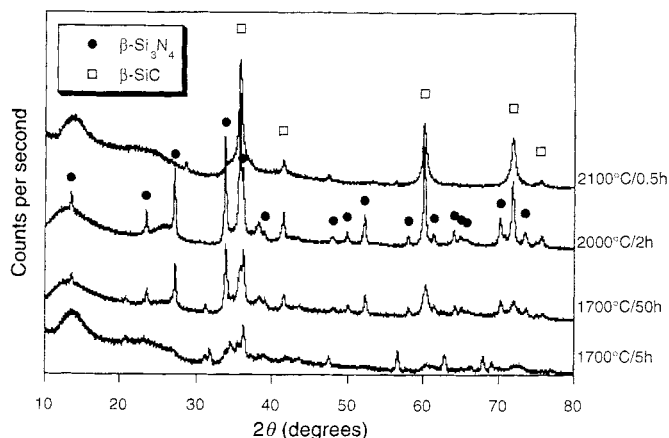


FIG. 3 X-ray powder diffraction (monochromatic CuK_α radiation with $\lambda = 154.178$ pm) of Si_{3.0}B_{1.0}C_{4.3}N_{2.0} isothermally annealed in the temperature range between 1,700–2,100 °C under 0.1 MPa Ar.

tensile strength is retained at those temperatures, as recently reported for Si–B–N–O- (ref. 20) and SiBN₃C-fibres (refs 14, 15). On the other hand, the *in situ* crystallization of single-phase and amorphous polymer-derived carbides and nitrides by annealing at 1,700 °C offers the possibility to synthesize multiphase and nanostructured ceramics from Si₃N₄ and SiC which, for thermodynamic reasons, cannot be produced by traditional methods^{9,10,21}. Polycrystalline composite ceramics comprised of two or more binary carbides or nitrides are presently of great interest with respect to design materials with tailor-made properties².

The extraordinarily high thermal stability found for silicoboron carbonitride enables the application of silicon-nitride-containing ceramics at temperatures exceeding 1,500 °C, which has been considered⁵ as the ultimate temperature for Si₃N₄. This new material therefore has considerable potential in technologies such as energy-efficient power generation, and mechanical and chemical engineering. The origin of the enhanced thermal stability is not yet understood, although our results suggest that boron plays a major role in stabilizing the metastable amorphous phase and in suppressing the thermal degradation of the silicon carbonitride matrix. □

Received 19 March; accepted 2 July 1996.

- Mitomo, M. & Tajima, Y. *J. Ceram. Soc. Japan* **99**, 1014–1025 (1991).
- Nihara, K. *J. Ceram. Soc. Japan* **99**, 974–982 (1991).
- Wakai, F. et al. *Nature* **344**, 421–423 (1990).
- Batha, H. D. & Whitney, E. D. *J. Am. Ceram. Soc.* **56**, 365–373 (1973).
- Raj, R. *J. Am. Ceram. Soc.* **76**, 2147–2174 (1993).
- Riedel, R., Seher, M., Mayer, J. & Szabo, D. V. *J. Eur. Ceram. Soc.* **15**, 703–716 (1995).
- Soderquist, J. A., Rivera, I. & Negron, A. *J. Org. Chem.* **54**, 4051–4055 (1989).
- Soderquist, J. A. & Lee, S.-J. *Tetrahedron* **44**, 4033–4042 (1988).
- Riedel, R., Kleebe, H.-J., Schönfelder, H. & Aldinger, F. *Nature* **374**, 526–528 (1995).
- Riedel, R., Passing, G., Schönfelder, H. & Brook, R. J. *Nature* **355**, 714–716 (1992).
- Takamizawa, M., Kobayashi, T., Hayashida, A. & Takeda, Y. US Patent No. 4604367 (1986).
- Baldus, H. P., Wagner, O. & Jansen, M. *Mater. Res. Soc. Symp. Proc.* **271**, 821–826 (1992).
- Baldus, H. P., Jansen, M. & Wagner, O. in *Silicon Nitride '93* (eds Hoffmann, M. J., Becher, P. F. & Petzow, G.) 75–79 (Trans Tech, Aedermannsdorf, 1994).

- Baldus, H. P. & Passing, G. *Mater. Res. Soc. Symp. Proc.* **346**, 617–622 (1994).
- Baldus, H. P., Passing, G., Sporn, D. & Thierauf, A. in *High-Temperature Ceramic-Matrix Composites II: Manufacturing and Materials Development* (eds Evans, A. G. & Naslain, R.) 75–84 (Ceramic Trans. 58, Am. Ceramic Soc., Westerville, Ohio, 1995).
- Löffelholz, J. & Jansen, M. *Adv. Mater.* **7**, 289–292 (1995).
- Funayama, O., Kato, T., Tashiro, Y. & Isodo, T. *J. Am. Ceram. Soc.* **76**, 717–723 (1993).
- Riedel, R., Kienzie, A., Petzow, G., Brück, M. & Vaahs, T. German Patent DE 4320784 A1 (1994).
- Becher, P. F. *J. Am. Ceram. Soc.* **74**, 255–269 (1991).
- Funayama, O., Nakahara, H., Tezuka, A., Ishii, T. & Isoda, T. *J. Mat. Sci.* **29**, 2238–2244 (1994).
- Mayer, J., Szabo, D. V., Rühle, M., Seher, M. & Riedel, R. *J. Eur. Ceram. Soc.* **15**, 717–727 (1995).

ACKNOWLEDGEMENTS. This work was supported by the Keramikverbund Karlsruhe-Stuttgart, the Deutsche Forschungsgemeinschaft, Bonn, and the Fonds der Chemischen Industrie, Frankfurt.

CORRESPONDENCE should be addressed to R.R. (e-mail: dg9b@hrzpub.th-darmstadt.de).

Fluid-mediated influence of adjacent thrusting on the seismic cycle at Parkfield

Stephen A. Miller

Geological Institute, Swiss Federal Institute of Technology (ETH),
CH-8092 Zurich, Switzerland

THE San Andreas fault near Parkfield, California has experienced moderate to large earthquakes about every 20 years since 1881, the most recent of which occurred in 1966¹⁻³. Statistical recurrence and precursory quiescence models^{4,5} led to a prediction of the next event for 1992 or before, but this event has not yet occurred. Here I present a model⁶ that attributes the 'non-occurrence' of the expected earthquake to the influence of adjacent thrusting on the rate of fault weakening from increasing pore pressures within the fault zone. Reductions in fault-normal strain rates⁷⁻⁹, and the onset of seismic quiescence along the San Andreas fault, both coincide with adjacent blind thrust faulting in the 1982-85 earthquake sequence of New Idria, Coalinga and Kettleman Hills. These events 'turned off' a compaction-induced weakening mechanism and extended the (proposed) seismic cycle at Parkfield. A model simulation of this effect shows a strong correlation in space and time of calculated and observed earthquake hypocentres, and shows that observed seismicity patterns from 1969-95 can be explained by reduced compaction rates brought on by adjacent thrusting. The model may have applications to other linked thrust/strike-slip fault systems, such as are found in southern California and other transpressional regions.

For this study, the area around Parkfield, California (Fig. 1) is divided into the Coalinga-Kettleman Hills (CK) region¹⁰, with thrust fault focal mechanisms, and the right-lateral strike-slip San Andreas fault (SAF) system. Near Parkfield, 478 seismic events of magnitude $M \geq 2.5$ were recorded by the Northern California Earthquake Data Center (NCEDC) between January 1969 and October 1995. The $M \geq 2.5$ cutoff was chosen to eliminate possible systematic errors in the catalogue due to increased detection capabilities of the network¹¹. The chosen region is ± 30 km along strike from Parkfield, from 0 to 15 km in depth, and ± 5 km from the SAF strike. All hypocentres within this range are projected onto a single fault plane for the analysis.

Dislocation models of the 1983 (moment magnitude, $M_w = 6.7$) Coalinga earthquake show changes in the normal and driving stress acting on the SAF (refs 12-14), but associated maximum stress changes of 0.1-0.2 MPa were not considered large enough to have long-term effects on the Parkfield region. However, the time history of seismicity for the CK region (Fig. 2a) and for the SAF region around Parkfield (Fig. 2b) clearly indicate a relationship between the increase of seismicity around Coalinga and the onset of quiescence around Parkfield. Here I argue that this is a causal relationship, attributable to changes in the rate of compaction-induced fault weakening within the SAF. The quiescence signal is found for all magnitude bands with $M \geq 2.0$, which satisfies the condition for real quiescence¹⁵ as opposed to artificial quiescence from systematic errors that are sometimes introduced into earthquake catalogues. The catalogue used in this study is not modified for possible magnitude shifts in the data. An analysis of trilateration data⁷ for 1966-84 showed compressional fault-normal strain rates of about $(-0.07 \pm 0.03) \times 10^{-6} \text{ yr}^{-1}$. Two-colour geodimeter measurements (started in 1984) recorded an extensional step of about 1×10^{-6} , coinciding with the 1985 Kettleman Hills earthquake⁹, followed by extensional strain rates of about $(0.15 \pm 0.01) \times 10^{-6} \text{ yr}^{-1}$ from 1986 to 1988. These observations imply that reduced fault-normal strain rates resulted from adjacent thrusting, but they are inconclusive without

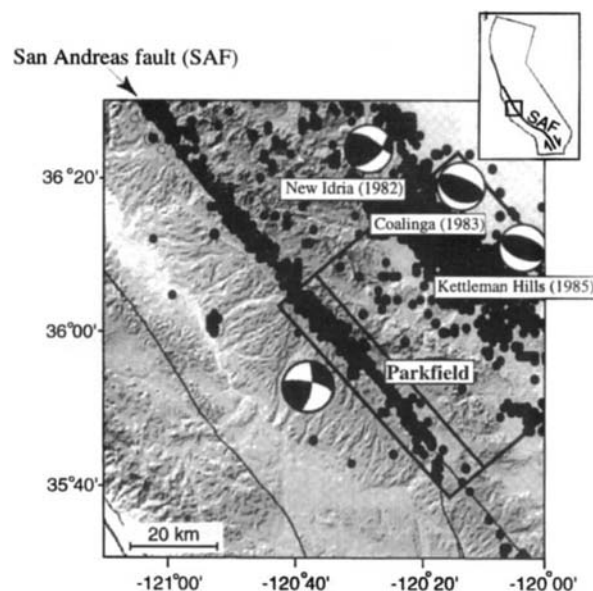


FIG. 1 Map of the study area (courtesy of US Geological Survey, USGS) with the epicentre (circles) of all events ($M \geq 2.5$) from January 1969 to October 1995 (data provided by the NCEDC). Concentrated epicentres mark the San Andreas fault, and the New Idria, Coalinga and Kettleman Hills regions are shown by the epicentres to the northeast. The beach balls show strike slip earthquake focal mechanisms along the San Andreas fault, and thrust fault focal mechanisms in the Coalinga region¹⁴. For results presented in this Letter, the region is separated into Coalinga-Kettleman Hills (CK), and 60 km along strike, centred around Parkfield (denoted by boxes).

additional analysis of the geodetic data^{8,16}. Changes in fault-normal strain rates influence the rate of pore pressure increase within the fault, which according to this model is the dominant mechanism driving seismicity in the area^{17,18}.

Recent models^{6,19-23} attribute the weakness of some faults to increased pore pressures within the fault zone. Earthquake clustering and repeated events can be explained by the movement of fluids^{6,17,18}. The model used in this study⁶ is a simple coupled system of pore pressure and shear stress, both of which are time varying. At present, I neglect other possible contributions from lithological, geometric and geochemical processes, and focus only on systems controlled by pore pressure. In the model, fault-normal compaction increases pore pressure (at rates dependent on the porosity reduction rate and pore compressibility) in zero-permeability pockets until a region slips by frictional sliding. At cell failure, the shear stress is redistributed within the fault plane using the solution of a rectangular dislocation in a three-dimensional elastic half-space²⁴⁻²⁶. Slip also allows fluid flow to neighbouring cells, and pore pressures are redistributed by conserving fluid mass. The decrease in pore pressure accompanying a slip event heals the failed cell by increasing the effective stress; all cells can fail, heal and fail again. This dynamical system between shear stress and effective stress is shown to self-organize²⁷ from a uniform shear stress on a strong fault to a strongly heterogeneous shear stress distribution on a weak fault. The model produces a nonlinear evolution to a weak fault⁶, and a power-law distribution of size with frequency.

The initial assumptions and input parameters were kept simple to demonstrate the general behaviour of this system and because the material properties at depth are not well constrained. I assumed simple friction (friction coefficient of 0.6), impermeable cell walls when cells are below the failure condition, but fluid communication between neighbouring cells when a cell slips. Material heterogeneity is imposed by a random uniform distribution of porosity and pore compressibility in a discretized matrix of 1,200 rectangular cells, 48 along strike and 25 through the depth, for a modelled fault measuring 60×15 km.

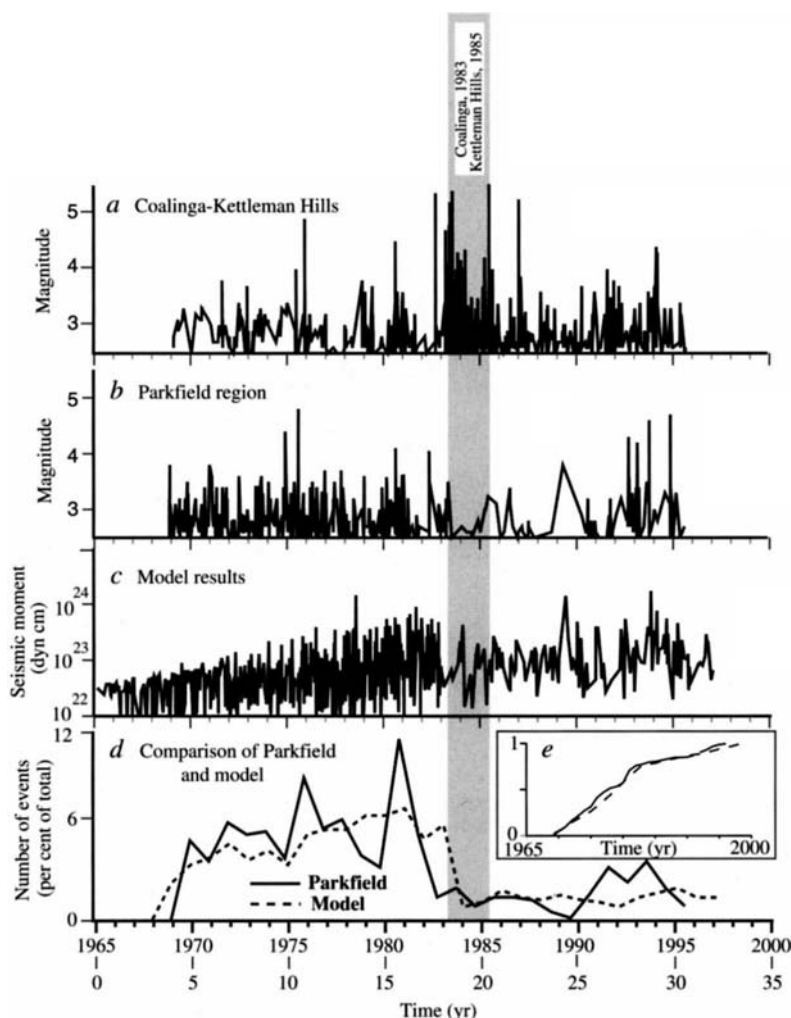


FIG. 2 Time history of seismicity for Coalinga-Kettleman Hills (a), Parkfield (b) and model results (c). A comparison of Parkfield seismicity and model results is shown in d as a percentage of the total number of events for Parkfield (478 events) and in the model (657 events). A comparison of the cumulative number of events (derived by integrating the distribution in d) is shown in e. The spatial distribution of records in b and c are compared in Figs 3 and 4.

To check the model against observed seismicity patterns and to test the hypothesis that adjacent thrusting delayed the expected Parkfield earthquake, the fault compaction rate was held constant until the effects of CK thrusting and their aftershocks were included by simply reducing the compaction rates. Modelled porosity reduction rates before CK thrusting are within an order of magnitude of measured closure rates along this part of the SAF (refs 7, 16), and range along strike from $6 \times 10^{-14} \text{ s}^{-1}$ at the northern end to $2 \times 10^{-14} \text{ s}^{-1}$ at the southern end of the fault. Changes in compaction rates along strike are assumed to occur where the fault-parallel ridge axis splays from the strike of the SAF towards the south, which indicates a decreasing fault-normal compression regime. To simulate adjacent thrusting, the porosity reduction rates were reduced to an assumed linearly increasing function from zero at the northern end to $1.5 \times 10^{-14} \text{ s}^{-1}$ at the southern end. The reduction in fault compaction rates is derived from the observed contraction before CK thrusting followed by extension after thrusting^{7,9}, and from dislocation analyses that show that the effects of the Coalinga earthquake, on fault-normal stresses, vary along strike^{13,14}. Pore pressure increase from a porosity reduction mechanism is assumed to operate below a depth of 3 km. The pore pressure buildup rates within the fault zone, before and after adjacent thrusting, associated with these input parameters are shown in Fig. 3. The initial stress state is

hydrostatic pore pressure, lithostatic mean stress and a depth-dependent shear stress distribution of 1 MPa km^{-1} . The shear stress is increased everywhere at a rate of 0.15 MPa yr^{-1} to include shear stress accumulation from plate motion. The model was conditioned for five simulated years to allow organization of the initial uniform stress field, so the model begins at $t = -5$ years, with $t = 0$ chosen as the 1966 Parkfield earthquake. I make no assumptions about a proposed recurrence interval along this stretch of the SAF.

The observed time history of seismicity around Parkfield (Fig. 2b) is compared with the calculated time history (Fig. 2c) in Fig. 2d. Figure 2d shows, as a percentage of the total number of events, how the observed and calculated events are temporally distributed. Both the observed and calculated results show alternating increases and decreases in the number of events up to 1983, followed by a period of relative quiescence until about 1990, when increased seismicity resumes. The observed and calculated cumulative number of earthquakes (derived by integrating the distributions in Fig. 2d) is shown in Fig. 2e. The temporal distribution of seismicity from Parkfield does not change with the choice of minimum magnitude (that is, the data is shown for $M \geq 2.5$, but the same shape is found for $M \geq 2.0, 2.1, 2.2, \dots, 3$ and so on), indicating a general behaviour of this system that is explained by a mechanism of dilatancy-accompanied slip^{21,22,28,29}. Quiescence follows a period of high activity because reduced fluid pressures which accompany slip heal the fault locally, and so additional time is required before another failure. This effect is captured by the model, but is damped because dilatancy-accompanied slip is not yet explicitly modelled. Conversely, quiescence precedes large events⁴ because the shear stress within the system reorganizes and is ultimately concentrated onto an evolved high-strength patch, which then loads quietly to failure.

A comparison of how the observed and calculated hypocentres are distributed along the fault plane (in three year increments) is shown in Fig. 3. These data correspond to the time histories in Fig. 2b and c. A qualitative agreement in the spatial and temporal hypocentre signature is observed. Specifically, both the data and model results show an evolution from diffuse seismicity to clustering and repeated events until 1983, followed by a period of quiescence and diffuse seismicity before evolving again to some clustering. Seismicity clustering returns, although the compaction rates are much lower, because accumulating shear stresses from plate motion are superposed on the complex shear stress distribution evolved from the initial state. The model also explains that the 'false alarm' recorded at Parkfield³ in 1992 could not have triggered the expected earthquake because insufficient reorganization of the shear stresses within the fault had occurred.

To quantify the distribution of hypocentres in the model and at Parkfield (Fig. 3), the time evolution of seismicity as a function of depth (Fig. 4a) was compared using a running average of 30 consecutive events³⁰. The spatial distribution of seismicity as a function of depth (as a percentage of the total number of events), is shown in Fig. 4b.

The overall agreement between model results and observations in both space and time suggests that the model captures the dominant mechanisms operating within this region of the SAF system. Adding complexity to the model (for example, slip weak-

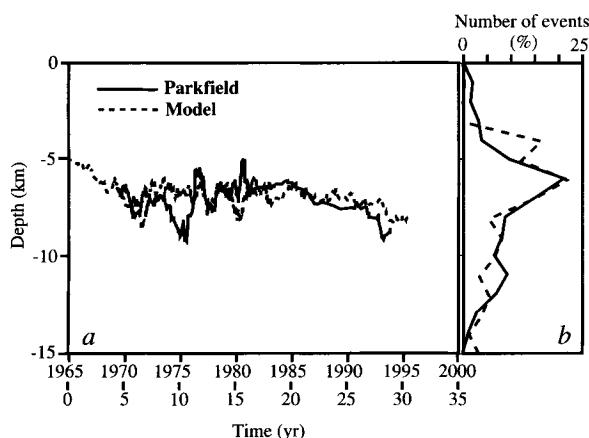
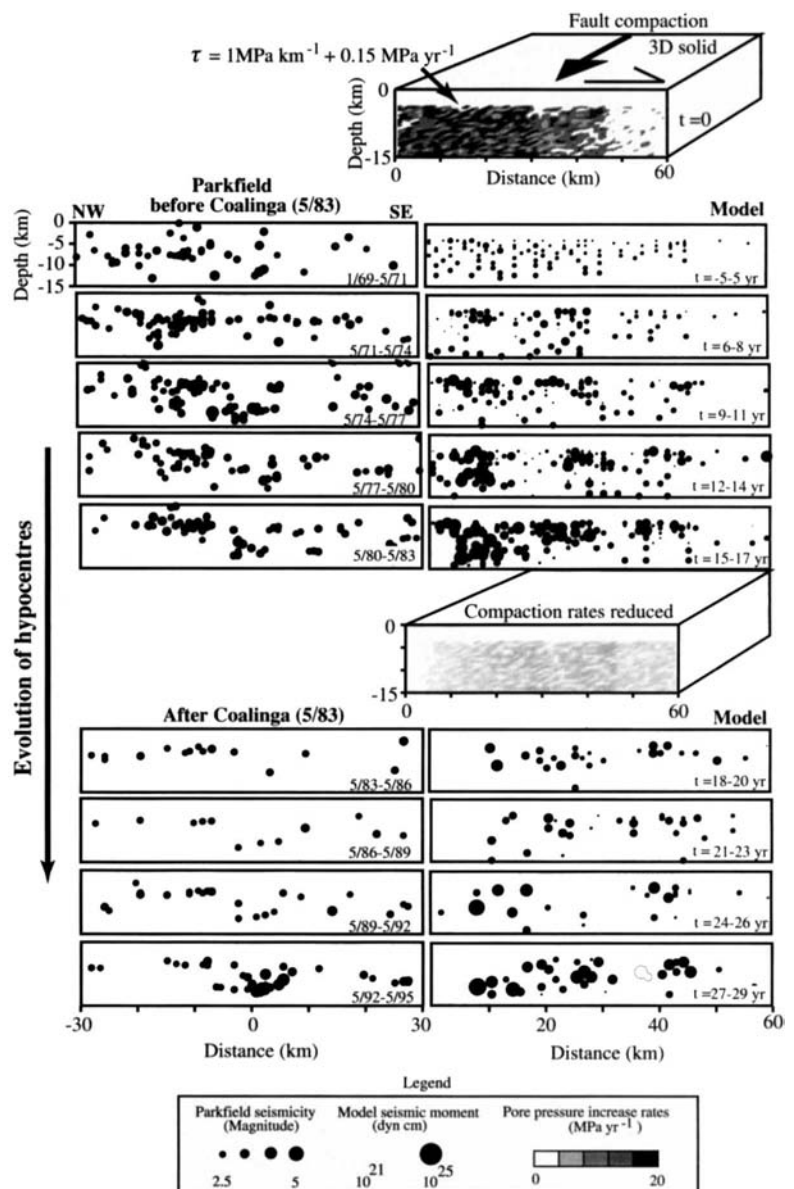


FIG. 4 *a*, Comparison of model and observations of the time evolution of the running averages of hypocentral depths of 30 consecutive events. *b*, Comparison of model and observations of the seismicity distribution as a function of depth.

ening, permeability–porosity relationships, dilatancy–slip and so on) may refine the correlation, but these are possibly second-order effects. The available data strongly suggest that the Parkfield earthquake did not happen as expected because Coalinga and associated thrusting did. The reduction in fault compaction rates brought on by adjacent thrusting suppressed a high-pore-pressure fault-weakening mechanism operating within the SAF. This allowed the self-organized, heterogeneous shear stress distributions developed around Parkfield to homogenize, so additional time was required to reorganize to a state for a large earthquake event. Changes in the long-term creep rates after 1983 are reconciled by the model because of the strengthening effect of homogenizing the shear stress and effective stress state along the fault. The model calls for a change in fault-normal strain rates associated with adjacent thrusting, but such a change must still be verified from the geodetic data. These results show that a periodic or semi-periodic seismic recurrence interval is not an appropriate model for short-term earthquake prediction if it ignores the mechanical effects of off-fault or far-field tectonic activity that influences future seismicity.

Forward modelling of the simulation presented shows a significant event in model year 34 (around year 2000), with a calculated hypocentre at 25 km and a depth of 9 km along the model fault. This is preceded by notable events at 21 and 24 km along the model fault, with hypocentral depths of 5 to 7 km. Additional mechanisms not considered here³¹ could lead to deviations from these expectations. A more detailed description of potential predictive capabilities of this model is beyond the scope of this Letter. □

Received 5 February; accepted 4 July 1996.

1. Bakun, W. H. & Lindh, A. G. *Science* **229**, 619–624 (1985).
2. Bakun, W. H. & McEvilly, T. V. *J. Geophys. Res.* **89**, 3051–3058 (1984).
3. Roeloffs, E. & Langbein, J. *Rev. Geophys.* **32**, 315–336 (1994).
4. Wyss, M., Bodin, P. & Habermann, R. E. *Nature* **345**, 428–431 (1990).
5. Savage, J. C. *Bull. Seismol. Soc. Am.* **83**, 1–6 (1993).
6. Miller, S. A., Nur, A. & Olgaard, D. L. *Geophys. Res. Lett.* **23**, 197–200 (1996).
7. Harris, R. A. & Segall, P. J. *Geophys. Res.* **92**, 7945–7962 (1987).
8. Dong, D. thesis, Massachusetts Inst. Technol. (1993).
9. Langbein, J. O., Burford, R. O. & Slater, L. E. *J. Geophys. Res.* **95**, 2533–2552 (1990).
10. Stein, R. S. & Ekström, G. *J. Geophys. Res.* **97**, 4865–4883 (1992).
11. Habermann, R. E. *Bull. Seismol. Soc. Am.* **77**, 141–159 (1987).
12. Poley, C. M., Lindh, A. G., Bakun, W. H. & Schulz, S. S. *Nature* **327**, 134–137 (1987).

13. Mavko, G. M., Schulz, S. & Brown, B. D. *Bull. Seismol. Soc. Am.* **75**, 475–489 (1985).
14. Simpson, R. W., Schulz, S. S., Dietz, L. D. & Burford, R. O. *Pure Appl. Geophys.* **126**, 665–685 (1988).
15. Wyss, M., Westerhaus, M., Berckhemer, H. & Ates, R. *Geophys. J. Int.* **123**, 117–124 (1995).
16. King, N. E., Segall, P. & Prescott, W. J. *Geophys. Res.* **92**, 2747–2766 (1987).
17. Johnson, P. A. & McEvilly, T. V. *J. Geophys. Res.* **100**, 12937–12950 (1995).
18. Nadeau, R. M., Foxall, W. & McEvilly, T. V. *Science* **267**, 503–507 (1995).
19. Rice, J. R. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-f.) 476–503 (Academic, San Diego, 1992).
20. Lockner, D. A. & Byerlee, J. D. *Pure appl. Geophys.* **145**, 717–745 (1995).
21. Sleep, N. H. & Blanpied, M. L. *Nature* **359**, 687–692 (1992).
22. Segall, P. & Rice, J. R. *J. Geophys. Res.* **100**, 22155–22171 (1995).
23. Byerlee, J. D. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
24. Ben-Zion, Y. & Rice, J. R. *J. Geophys. Res.* **98**, 14109–14131 (1993).
25. Stuart, W. D., Archuleta, R. J. & Lindh, A. G. *J. Geophys. Res.* **90**, 592–604 (1985).
26. Chinnery, M. A. *Bull. Seismol. Soc. Am.* **53**, 921–932 (1963).
27. Bak, P. & Tang, C. J. *Geophys. Res.* **94**, 15635–15637 (1989).
28. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
29. Sibson, R. H. *Earth Sci. Rev.* **32**, 141–144 (1992).
30. Wyss, M. *Geophys. Res. Lett.* **18**, 617–621 (1991).
31. Ben-Zion, Y., Rice, J. R. & Dmowska, R. *J. Geophys. Res.* **98**, 2135–2144 (1993).

ACKNOWLEDGEMENTS. I thank C. H. Scholz for encouragement and discussions; E. Roeloffs for a review; and T. Meckel, J.-P. Burg, M. Häni, and N. Mantelov for improvements to the manuscript. Special thanks and appreciation to A. Nur and D. L. Olgaard.

CORRESPONDENCE and requests for materials should be addressed to S.A.M. (e-mail: steve@erdw.ethz.ch).

Flavodoxin as an *in situ* marker for iron stress in phytoplankton

Julie La Roche*, Philip W. Boyd†, R. Michael L. McKay*‡ & Richard J. Geider§

* Oceanographic and Atmospheric Sciences Division, Brookhaven National Laboratory, Upton, New York 11973-5000, USA
 † Department of Oceanography, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada
 ‡ College of Marine Studies, University of Delaware, Lewes, Delaware 19958, USA
 § Marine Biological Association of the UK, Plymouth Marine Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

A fundamental issue in marine science is the identification of the factors controlling biological uptake of CO₂ in high-nitrate, low-chlorophyll regions^{1–10}. A recent *in situ* iron fertilization experiment demonstrated that iron limitation is responsible for low phytoplankton stocks in the equatorial Pacific⁴. Here we show that flavodoxin, a biochemical marker of iron limitation, can be used to map the degree of iron stress in natural populations. Flavodoxin assays along a 900-km east-west transect in the northeastern subarctic Pacific revealed a pronounced increase in iron stress in the region west of the 135° W meridian. Addition of dissolved iron alleviated this stress. Immunostaining of single cells from the most western station showed that flavodoxin is present specifically within the chloroplasts of diatoms. Our approach provides a rapid means of defining the extent of iron stress in the ocean⁵ and supports the hypothesis that diatoms are iron stressed in the northeast Pacific^{6,10}.

Flavodoxin is a redox protein that contains a single molecule of riboflavin 5'-phosphate as a cofactor. Under iron-stress conditions, the iron-sulphur protein ferredoxin is replaced by flavodoxin in many microorganisms¹¹ including diatoms^{12,13}. Because iron is an essential component of several proteins involved in photosynthetic and respiratory electron transport, flavodoxin expression at iron concentrations that are suboptimal for growth is expected to partially alleviate iron stress¹³. We have previously shown that the cellular content of flavodoxin is inversely correlated with the initial amount of iron in the culture medium¹³. Both flavodoxin and the ratio of flavodoxin/ferredoxin have been correlated with algal growth rates in laboratory cultures^{13,14}. We measured flavodoxin abundance relative to total protein in extracts from phytoplankton samples along an east-west transect

from the British Columbia coast (Fig. 1a; P04) to the former site of ocean station Papa (P26) in May and September 1995 using polyclonal antisera raised against flavodoxin that had been isolated from the diatom *Phaeodactylum tricornutum*. This antiserum cross-reacts with high affinity to all diatom species examined to date ($n = 11$) but not significantly with other taxa¹³.

Dissolved iron and nitrate concentrations were inversely correlated along the transect, with the highest concentrations of iron observed at the nearshore station (P04) and the highest nitrate concentrations observed at the offshore station (P26) (Fig. 1). Flavodoxin levels were near the detection limit at P04, the nearshore station, and more than 30 times higher at stations P20 and P26 (Fig. 2). The trends in flavodoxin abundance were reproducible in both May and September 1995, although the phytoplankton taxonomic composition varied widely along the transect and between cruises (see Fig. 1 legend). The similar increases in flavodoxin levels from east to west on the two cruises, despite the shift in the composition of the phytoplankton assemblages, provides evidence for a persistent physiological response to iron stress. Although we have not yet established a relationship between flavodoxin and growth rate in the field, flavodoxin levels remained low or undetectable (<10% of the levels measured at P26) in the presence of dissolved iron concentrations greater than 1 nM, the concentration above which algal growth rates become iron saturated in the equatorial Pacific¹⁵.

In general, flavodoxin expression in microorganisms is regulated by iron availability. However, there is taxonomic variability ranging from constitutive expression to a complete absence of flavodoxin¹¹. Verification that cellular flavodoxin is regulated by dissolved iron is needed, especially in oceanic regions which have chronically low levels of this micronutrient. Iron-addition experiments conducted at station P26 demonstrate that flavodoxin abundance is regulated by iron availability (Fig. 3), with transient reductions in flavodoxin levels observed in the iron-amended samples. The residual flavodoxin signal remaining after iron addition indicates that the cells attained only partial alleviation of iron stress. A constant supply of iron may be necessary to alleviate iron stress completely^{16,17}. Sustained addition, rather than a discrete addition of a relatively high concentration of iron may alleviate iron stress for a longer period⁴. Little is known about the turnover time of flavodoxin, and the residual signal in flavodoxin could alternatively reflect the biological half-life of previously synthesized flavodoxin. Increased flavodoxin abundance in the controls demonstrates that the enclosure of a natural seawater sample in a bottle generates additional iron stress. It is likely that the increase in flavodoxin:protein ratio in the controls is a consequence of uptake of the residual

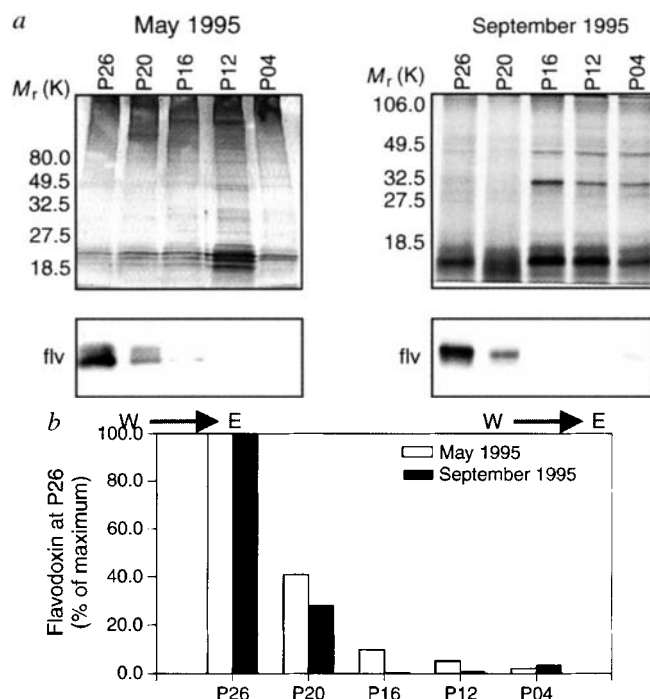
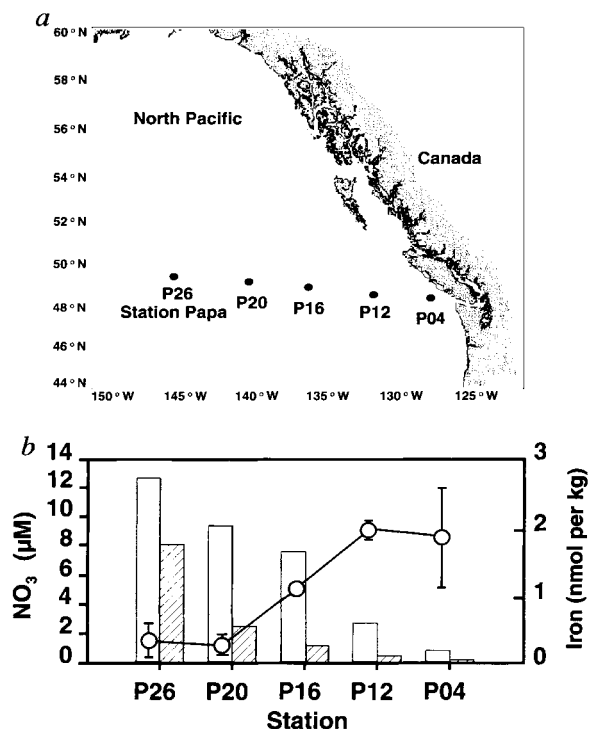
bioavailable iron by phytoplankton and subsequent decrease in the cellular concentration of metabolic iron. During September, macronutrient-addition experiments at the near-shore station (P04) were used to deplete dissolved iron relative to other macronutrients. Induction of flavodoxin after large additions of iron-free macronutrients to samples from this relatively iron-rich region demonstrates that flavodoxin is iron regulated in endemic diatom populations at P04 (data not shown). Similar results were reported previously for eastern US coastal waters characterized by high dissolved iron levels and undetectable *in situ* flavodoxin levels¹³.

FIG. 1 a, Map of the northeastern subarctic Pacific showing stations P04 (1,000 m depth)—P26 (4,200 m depth) along the transect. b, Dissolved iron concentrations (circles) and nitrate levels (bars: white, May 1995; hatched, September 1995) in the surface mixed layer (0–40 m). Dissolved iron²⁴ (mean of all available measurements $\pm$ s.d., except for P04 where the error bar is the range of two observations) and nitrate were measured in samples obtained using clean techniques¹⁰. The mean dissolved iron value at station P26 is based on three Canadian Joint Global Ocean Flux Study cruises and one historical measurement²⁵. Dissolved iron decreased by roughly one order of magnitude between inshore (P04) and offshore stations (P20 and P26). Our dissolved iron levels at P26 are up to 10 times higher than those previously reported²⁵ and may have resulted from shipboard contamination¹⁰. Chlorophyll concentrations were uniform during May 1995 (mean, $0.34 \mu\text{g l}^{-1}$; s.d., $0.05 \mu\text{g l}^{-1}$) and were lower at stations P04, P12 and P16 ($0.1\text{--}0.14 \mu\text{g l}^{-1}$) than at stations P20 and P26 (0.31 and $0.36 \mu\text{g l}^{-1}$, respectively) during September 1995. The high concentrations of chlorophyll at P20 and P26 in September were due to high dinoflagellate and diatom abundances. Autotrophic nanoflagellates were numerically the most abundant phytoplankton group at all stations, ranging from $1.2 \times 10^5 \text{ l}^{-1}$ at P12 in May 1995 to $31 \times 10^5 \text{ cells l}^{-1}$ at P20 in September 1995. Diatoms were present at 79, 34, 23, <1 and $5.5 \times 10^4 \text{ cells l}^{-1}$ during May 1995 at stations P26, 20, 16, 12 and P04, respectively. During September 1995, diatoms were present at 8.3, 27, 1.3, 0.9, $11 \times 10^4 \text{ cells l}^{-1}$ during May 1995 at stations P26, 20, 16 and P04, respectively. Autotrophic nanoflagellate and diatom numbers at P26 concur with previous observations¹⁸. Diatom biomass, which is targeted by the bulk immunological assay, was higher at the inshore station (P04) than at P26 during the May cruise. During September, diatom biomass was greater at P20 because of the high abundance of several large-sized diatom species not normally observed at this station (P.W.B., unpublished results). With the exception of P04 (May and September) and P20 (September only), diatoms were mainly small (< 10 μm) pennates.

FIG. 2 Flavodoxin abundance along the transect from P26 to P04 for May and September 1995. a, Silver-stained gels (top panel) and western blots treated with anti-flavodoxin (flv) (bottom panel) showing that flavodoxin levels were highest at station P26 and declined from west (W) to east (E) along the transect. b, Flavodoxin levels expressed as per cent of the maximum amount found at P26. Flavodoxin levels at station P26 during the September cruise were >3 times higher than at station P20, and >150-times higher than at stations P16 and P12. Similar results were obtained when flavodoxin levels were normalized to Rubisco levels determined immunologically (results not shown).

METHODS. Water (60 l) was collected from the mixed layer using 30 l Go-flo samplers¹⁰ and transferred via tubing directly to a filtration apparatus under slight positive pressure. Phytoplankton were collected on polycarbonate filters (142 mm diameter, 5 μm porosity). The cells were rinsed from the filter and concentrated into a pellet by centrifugation. Cell pellets were stored in liquid nitrogen before analysis on shore. Total protein was extracted by sonicating cell pellets in acetone containing 10% (w/v) trichloroacetic acid and 0.07% (v/v) 2-mercaptoethanol. Protein yield was 50–200 μg per 60 l sample. Trichloroacetic acid precipitable protein was resuspended in reducing sample buffer¹³ and equal amounts of protein (10 μg) per lane were separated by SDS-PAGE on 4–20% gradient gels (May 1995) or 15% acrylamide gels (September 1995). Gels were either silver stained²⁶ or electrophoretically transferred to nitrocellulose²⁷ (triplicate gels) for immunoreaction with polyclonal antiserum raised against diatom flavodoxin¹³. Immunoreactive proteins were detected by chemiluminescence on pre-flashed film and reaction products quantified by laser densitometry.

A single-cell immunoassay for flavodoxin (Fig. 4) was used on field samples to verify that the bulk assay reacted principally with diatoms. A strong cross-reactivity was observed with diatoms commonly found at P26 (refs 10, 18) but not with other taxa. This demonstrates that our immunoassay is specific for diatoms, which have been shown to respond most strongly to iron enrichments in high-nitrate, low-chlorophyll regions^{4,6,10}. It also demonstrates the potential for a single-cell immunoassay to distinguish differences in iron stress among phytoplankton taxa. Such an approach is useful given the wide variation in cellular iron quotas observed between species from the same taxonomic group¹⁹.



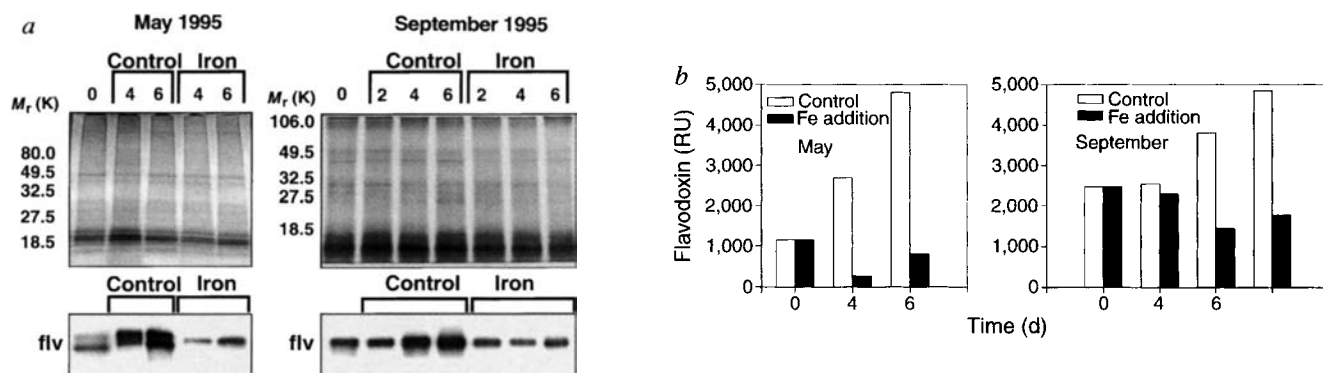


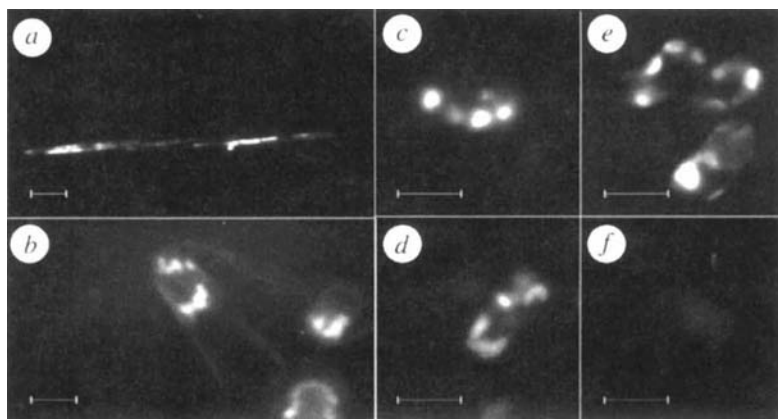
FIG. 3 Results of *in vitro* iron-enrichment experiments at station P26 in May and September 1995. **a**, Silver-stained gels (top panel) and western-blot analysis (lower panel) of samples incubated 4 and 6 (May 1995) or 2, 4 and 6 (September 1995) days after the initial treatment (0). **b**, Quantitative analysis of western blots for iron-addition experiments demonstrating that flavodoxin levels (RU, relative units) increased over the course of the experiment in control bottles and decreased transiently in the iron-amended bottles.

METHODS. Water was collected and transferred into clean 25-l polycarbonate carboys¹⁰. Controls contained unamended seawater, whereas dissolved iron (in a 1:1.5 iron:EDTA chelated complex) was added to

experimental bottles to a final concentration of 100 nM dissolved iron. After addition of iron, subsamples (8 l) were drawn after 2, 4 and 6 d. Protein isolation, electrophoresis and flavodoxin detection were as in Fig. 2. Residual nitrate (3.2 μM in May 1995 and 0.1 μM in September 1995) remained in the iron-amended bottles after 6 days of incubation. Changes in chlorophyll concentrations in controls and iron-amended samples were similar to those observed previously at this station^{10,25}. Tenfold increases in chlorophyll (0.35 to 3.45 $\mu\text{g l}^{-1}$ for September 1995) were generally observed in iron-amended bottles after six days. In contrast, chlorophyll in control bottles typically doubled (0.35 to 0.65 $\mu\text{g l}^{-1}$ for September 1995) during the same period.

FIG. 4 Single-cell immunofluorescence assay showing the presence of flavodoxin in diatoms from station P26. Images of **a**, pennate diatom and **b**, *Chaetoceros* sp. from the May and September cruises respectively, captured with a CCD camera system. As expected, staining of flavodoxin is intense and is confined to the chloroplast of each cell. Similar results were obtained from station P20 (results not shown). Controls with primary anti-serum omitted showed no specific staining. As a further control, *Thalassiosira weissflogii* cultured in the laboratory under iron-limiting (**c**, **e**) or iron-replete (**d**, **f**) conditions was probed for Rubisco (**c**, **d**) or flavodoxin (**e**, **f**). Whereas the chloroplasts of iron-limited cells are stained by anti-flavodoxin, no specific staining pattern is established in iron-limited cells. In contrast, chloroplast-specific staining is observed in all cells when treated with antiserum raised against chromophyte Rubisco holoenzyme.

METHODS. Cells were fixed in 100 mM sodium phosphate buffer (pH 7.2) containing 4% (w/v) freshly depolymerized paraformaldehyde for 6 h at 4 °C (ref. 28) and stored in methanol at -20 °C until assayed on shore. Cells affixed to poly-L-lysine-coated microscope slides were permeabilized with 0.05% (v/v) dimethyl sulphoxide followed by incubation with affinity-purified flavodoxin antiserum. Immunoreaction was



detected using fluorescein isothiocyanate-coupled secondary antiserum. All reagents were diluted in phosphate-buffered saline. Immunofluorescence was viewed with a Nikon Optiphot-2 microscope equipped with epifluorescence attachment and B-2A filter block. Scale bars, 10 μm .

which further implies that the threshold below which iron becomes limiting should vary between taxa.

The complex nature of the chemistry of iron in seawater and the importance of iron-supply rates to the physiological status of phytoplankton²⁰⁻²² indicate difficulties in using dissolved iron concentrations to infer biological iron stress. Our findings demonstrate that molecular markers, such as flavodoxin, can be used to probe the physiological status of unicellular marine organisms in natural populations²³. Further development of such markers for use in natural populations, in conjunction with bioassay experiments at selected sites, will allow routine monitoring of the physiological response to iron stress and the definition of the extent of such stress over large temporal and spatial scales in the open ocean. Our results are in agreement with the mounting evidence that iron supply controls diatom stocks in the northeastern subarctic Pacific^{6,10}. Further, the results of our surveys indicate that there are

marked differences in iron stress within this high-nitrate, low-chlorophyll region, such that cells to the east of the 135° W meridian seem to be iron replete. □

Received 5 May; accepted 5 July 1996.

1. Martin, J. H. *Paleoceanography* **5**, 1-13 (1990).
2. Cullen, J. J. *Limnol. Oceanogr.* **36**, 1578-1599 (1991).
3. Martin, J. H. et al. *Nature* **371**, 123-129 (1994).
4. Coale, K. et al. *Eos* **76**, OS154 (1996).
5. Longhurst, A. *Nature* **379**, 585-586 (1996).
6. Martin, J. H. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 123-137 (Plenum, New York, 1992).
7. Banse, J. G. R. *Limnol. Oceanogr.* **35**, 772-775 (1990).
8. Chisholm, S. W. & Morel, F. M. M. *Limnol. Oceanogr.* **36**, (1991).
9. de Baar, H. J. W. et al. *Nature* **373**, 412-415 (1995).
10. Boyd, P. W. et al. *Mar. Ecol. Prog. Ser.* **136**, 179-196 (1996).
11. Geider, R. J. & LaRoche, J. *Photosynth. Res.* **39**, 275-301 (1994).
12. LaRoche, J., Geider, R. J., Graziano, L. M., Murray, H. & Lewis, K. J. *Phycol.* **29**, 767-777 (1993).

13. LaRoche, J., Murray, H., Orellana, M. & Newton, J. J. *Phycol.* **31**, 520–530 (1995).
14. Doucette, G. J., Erdner, D. L., Peleato, M. L., Hartman, J. J. & Anderson, D. M. *Mar. Ecol. Prog. Ser.* **130**, 269–276 (1996).
15. Coale, K. H., Fitzwater, S. E., Gordon, R. M., Johnson, K. S. & Barber, R. T. *Nature* **379**, 621–624 (1996).
16. Kolber, Z. S. et al. *Nature* **371**, 145–149 (1994).
17. Behrenfeld, M. J., Bale, T., Aiken, J., Kolber, Z. S. & Falkowski, P. G. *Eos* **76**, OS176 (1996).
18. Booth, B. C., Lewin, J. & Postel, J. R. *Prog. Oceanogr.* **32**, 57–99 (1993).
19. Sunda, W. G. & Huntsman, S. A. *Mar. Chem.* **50**, 189–206 (1995).
20. Wells, M. L., Price, N. M. & Bruland, K. W. *Mar. Chem.* **48**, 157–182 (1995).
21. Barbeau, K., Moffett, J. W., Caron, D. A., Croot, P. L. & Erdner, D. L. *Nature* **380**, 61–64 (1996).
22. Wu, J. & Luther, G. W. *Mar. Chem.* **50**, 159–177 (1995).
23. Entsch, B., Sim, R. G. & Hatcher, D. G. *Mar. Biol.* **73**, 17–30 (1983).
24. Yang, L. thesis, Univ. British Columbia (1993).
25. Martin, J. H., Gordon, R. M., Fitzwater, S. & Broenkow, W. W. *Deep Sea Res.* **35**, 177–196 (1989).
26. Rabilloud, T., Carpentier, G. & Tarroux, P. *Electrophoresis* **9**, 288–291 (1988).
27. Towbin, H., Staehelin, T. & Gordon, J. *Proc. Natl Acad. Sci. USA* **76**, 4350–4354 (1979).
28. Lin, S., Chang, J. & Carpenter, E. J. *J. Phycol.* **31**, 388–395 (1995).

ACKNOWLEDGEMENTS. We thank J. Berges, A. Milligan, H. Murray-Fobin, F. Whitney, R. Walker and the officers and crew of the RV John P. Tully for their assistance, R. Chretien for determinations of DfE made on samples collected by P.W.B. and P. J. Falkowski for the Rubisco antiserum. D. Wallace, P. J. Harrison and T. Pedersen provided helpful comments and discussions. This research was supported by the NSF (R.J.G. and G. W. Luther), the US Department of Energy Office of Health and Environmental Research (J.L.), and by the NSERC (Canada) JGOFS programme (P.W.B. and P. J. Harrison).

CORRESPONDENCE and requests for materials should be addressed to J.L.R. (e-mail: Laroche@BNLUX1.BNL.GOV).

Two different areas within the primary motor cortex of man

Stefan Geyer*, Anders Ledberg†, Axel Schleicher‡, Shigeo Kinomura†, Thorsten Schormann‡, Uli Bürgel‡, Torkel Klingberg†, Jonas Larsson†, Karl Zilles*‡ & Per E. Roland†

* Department of Neuroanatomy and ‡ C. and O. Vogt Institute for Brain Research, Heinrich Heine University Düsseldorf, P. O. Box 101007, 40001 Düsseldorf, Germany

† Division of Human Brain Research, Department of Neuroscience, The Karolinska Institute, Doktorsringen 6F, 17177 Stockholm, Sweden

THE primary motor area (M1) of mammals has long been considered to be structurally and functionally homogeneous^{1–5}. This area corresponds to Brodmann's cytoarchitectural area 4. A few reports showing that arm and hand are doubly represented in M1 of macaque monkeys^{6,7} and perhaps man⁸, and that each subarea has separate connections from somatosensory areas, have, with a few exceptions^{9–12}, gone largely unnoticed. Here we show that area 4 in man can be subdivided into areas '4 anterior' (4a) and '4 posterior' (4p) on the basis of both quantitative cytoarchitecture and quantitative distributions of transmitter-binding sites. We also show by positron emission tomography that two representations of the fingers exist, one in area 4a and one in area 4p. Roughness discrimination activated area 4p significantly more than a control condition of self-generated movements. We therefore suggest that the primary motor area is subdivided on the basis of anatomy, neurochemistry and function.

We analysed sections of five human brains with quantitative cytoarchitectonic methods¹³. The border between somatosensory area 3a and the posterior part of area 4 was sharp. Large, slender pyramidal neurons in layer III characterized area 6, and the transition to the anterior part of area 4 was distinct. Differences in volume densities of neurons in layers II to VI between areas 3a and 4 as well as between 6 and 4 were statistically significant (Fig. 1). Our quantitative methods also revealed a statistically significant difference between the laminar density of neurons in the posterior part of area 4, area 4p and its anterior part, area 4a (Fig. 1). Area 4a contained slightly larger, more densely packed pyramidal cells in layer III than did 4p. Differences between 4a and 4p were also detected by *in vitro* receptor autoradiography. Both

regional and laminar densities of [³H]oxotremorine-M binding sites (muscarinic M₂ and M₄ receptors) were much higher in 4p than in 4a (Fig. 2a). There was a clear border between these areas which corresponded exactly to one marking differences in laminar densities of neurons in an adjacent Nissl section (Fig. 2b). Another

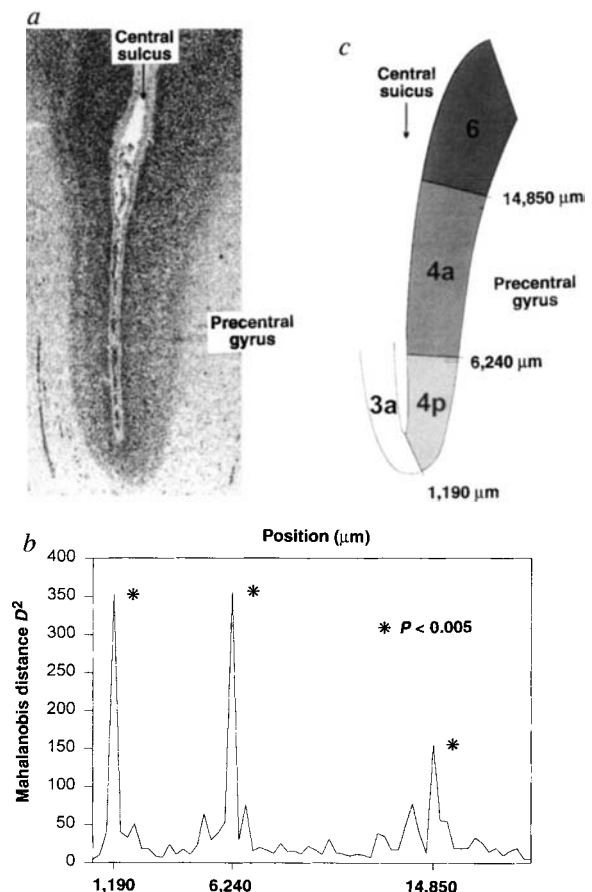


FIG. 1 Quantitative cytoarchitecture of area 4. *a*, Grey-level index image showing densities of neurons in a Nissl-stained section through pre- and postcentral gyrus, roughly half-way between interhemispheric and Sylvian fissures. *b*, Plot of Mahalanobis distance D^2 as a function of length measured along the middle of layer IV. At positions 1,190, 6,240 and 14,850 μm from the bottom of the sulcus, the cytoarchitecture changes significantly. These positions correspond to borders between areas 3a/4p, 4p/4a and 4a/6, as shown in *c*.

METHODS. Coronal 20- μm paraffin sections were stained with a modified Nissl method²¹. Volume density of cells was estimated with the grey-level index (GLI) procedure¹³. Each pixel represents neuronal density in a field measuring 27 μm per side. Density profiles, oriented orthogonally to the laminae and extending from the boundary between layers I and II to that between layer VI and white matter were extracted from the images. Profiles were standardized to a cortical depth of 100%. Each profile was 11 pixels (297 μm) wide, that is each value in a profile represents the average of 11 GLI values. Spacing between profiles was 297 μm . Characterizing features were extracted from each profile. They included the mean amplitude and the first 4 moments (mean, s.d., skewness, kurtosis). A mean feature vector $\bar{\mathbf{X}}_1$ was calculated from a block of 10 adjacent profiles, and another mean feature vector $\bar{\mathbf{X}}_2$ from a neighbouring block of 10 profiles. The Mahalanobis distance²² $D^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)' \mathbf{C}^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$ was calculated from the vectors and the inverse of the pooled covariance matrix $\mathbf{C}^{-1}$. Plotting the D^2 values as a function of the position of the blocks of profiles revealed maxima at locations at which regions covered by profiles showed large differences in laminar patterns. Statistical significance was established by computing Hotelling's statistic $T^2 = D^2 / (1/n_1 + 1/n_2)$ in which $n_1 = n_2 = 10$, the number of profiles from which each of the mean vectors had been calculated. The corresponding *P* value was not corrected for multiple comparisons, as the probability of getting three such maxima at the predicted locations in a set of adjoining sections is exceedingly small.

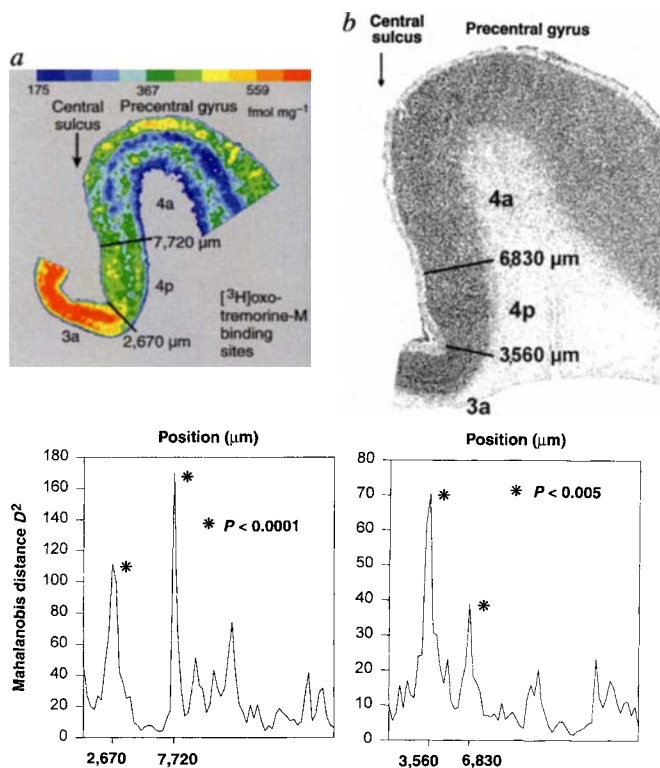
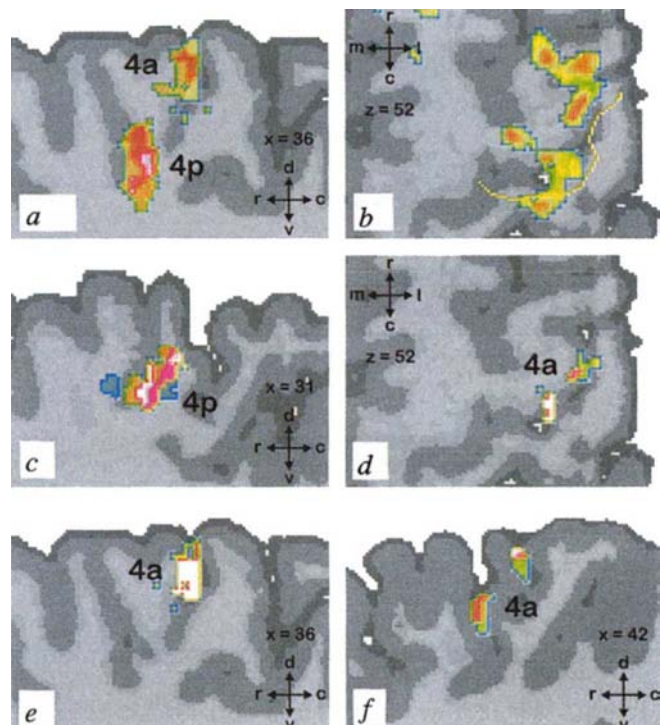


FIG. 2 *a*, Contrast-enhanced, colour-coded autoradiograph of section through precentral gyrus showing distribution of [^3H]oxotremorine-M binding sites on muscarinic M_2 and M_4 receptors (upper panel). B_{max} values expressed as fmol per mg protein. *b*, GLI image (see Fig. 1) of Nissl-stained adjacent section (upper panel). Observer-independent mapping revealed statistically significant maxima in Mahalanobis distance D^2 (ordinate; $P < 0.0001$ and $P < 0.005$) at positions 2,670 and 7,720 μm (abscissa) from bottom of sulcus in autoradiograph (*a*, lower panel) and at positions 3,560 and 6,830 μm in GLI plot (*b*, lower panel). These positions represent borders between areas 3a/4p and 4p/4a, respectively. Border between 4a and 6 not shown.

METHODS. Fresh, unfixed tissue was sectioned with a cryostat (20 μm). For GLI procedure, see Fig. 1. [^3H]oxotremorine-M binding sites (total binding) were labelled by quantitative *in vitro* receptor autoradiography²³. Autoradiographs were digitized with a CCD-camera. Standards of known radioactivity were used to compute a transformation curve indicating the relationship between grey values in the autoradiograph and concentrations of radioactivity in the tissue. Concentration values were multiplied by $(K_d + c)/c$ (K_d , equilibrium dissociation constant of ligand-binding kinetics = 1.74 nM; c , free concentration of labelled ligand = 0.86 nM) to obtain B_{max} values. Non-specific binding, determined by [^3H]oxotremorine-M labelling in the presence of 1 μM of unlabelled carbachol, was less than 5% of total binding. After calculating B_{max} values, autoradiographs were subjected to linear contrast enhancement, colour coding and median filtering (matrix size 5×5 pixels). Equidistant density profiles (297 μm wide), extending from the pial surface to the layer VI/white matter boundary, were extracted from the original data. They were oriented perpendicularly to lines drawn at these limits and standardized to a cortical depth of 100%. D^2 and significance of maxima were calculated as described for Fig. 1.

FIG. 3 Images of the standard brain with cytoarchitecture and functional activations in identical anatomical format¹⁴. *a*, Sagittal section 36 mm to the left of the midline ($x = 36$) showing those regions in which 4a and 4p were located in at least 3 of 5 brains. Yellow: overlap in 3 of 5 brains, red: 4 of 5, white: 5. *r*, rostral; *c*, caudal; *d*, dorsal; *v*, ventral. *b*, Horizontal section 52 mm above the midcommissural plane ($z = 52$), showing two activations in the left hemisphere associated with rapid flexions of the right thumb in the cortex lining the anterior wall of the central sulcus (somatosensory reaction-time task). The average course of the central sulcus is also shown. *m*, medial; *l*, lateral. *c*, Sagittal section ($x = 31$) showing in area 4p the overlapping activations of the two reaction-time tasks (thumb), roughness discrimination (thumb and index finger), and discrimination of ellipsoids (mainly thumb, index and middle finger). White: all tasks overlap; red: overlap of roughness discrimination and ellipsoid discrimination; yellow: overlap of ellipsoid discrimination and one of the reaction-time tasks. Centres of gravity (x, y, z): overlap of the two reaction-time tasks, 33, -27, 50; overlap of all tasks, 33, -28, 50. *d*, Horizontal section ($z = 52$) showing the overlap within area 4a (white; 38, -22, 55) of visual-memory task (index and middle finger) and ellipsoid discrimination (thumb, index and middle finger). *e*, Sagittal section ($x = 36$) showing the overlap in area 4a (white) of visual-memory task and ellipsoid discrimination. *f*, Sagittal section ($x = 42$) showing the overlap of activations of the two reaction-time tasks (thumb; lower region, 42, -13, 52) and another overlap between the two reaction time tasks (thumb) and ellipsoid discrimination (thumb, index and middle finger; upper region), both within area 4a. The localization of the overlaps within 4a and 4p were determined by Boolean operations on the images, thus *c-f* show only those parts of the overlaps which belonged to the space defined in *a*.

METHODS. Tasks: First group ($n = 10$): two reaction-time tasks (control state, complete rest²⁴); visual, key press with right thumb in response to a change in luminance of a yellow monochrome circle; somatosensory, key press at the protrusion of a stylus to the tip of the right index finger. Second group ($n = 10$): visual memory task; key press with right index or middle finger depending on which of two pictures belonged to a previously learned picture pair (visually matched control, no movement). Third group ($n = 9$): tactile discrimination of shapes of two ellipsoids with fingers of the right hand by a two-alternative forced-choice procedure¹⁸ (control, rest with eyes fixating a cross on the screen hiding the ellipsoids). Fourth group ($n = 9$): discrimination of roughness of two polyoxymethylene cylinders with different microprofiles with right thumb and index finger by a two-alternative forced-choice procedure¹⁸. All subjects and brains were scanned with a 1.5 T magnetic resonance (MR) scanner with 3D-spoiled gradient (3D-SPGR) or 3D-fast low angle shot (3D-FLASH) pulse sequences giving 1 mm³ isotropic voxels. For PET methods and treatment of data see refs. 17, 25. Scans of regional cerebral blood flow (rCBF) were



obtained with an in-plane spatial resolution of 4.5 mm and an interslice distance of 6.5 mm. Subjects received a bolus injection of 70 mCi of [^{15}O]butanol. Arterial tracer radioactivity was continuously monitored and rCBF values calculated on the basis of data collected during the first 60 s after injection²⁵. rCBF images were anatomically standardized as described above¹⁷. Final image resolution was 4.2 mm in plane²⁵. Subtraction images were made by voxel-by-voxel subtraction of control rCBF conditions from motor conditions. Only clusters of high t-value having a probability of less than 0.05 of being false positives within the space of the brain were considered statistically significant²⁵. The probability that such clusters should be localized to the space covered by areas 4a and 4p is then less than 0.0005.

corresponding border within Brodmann's area 4 separated areas with differing densities of [^3H]ketanserin-binding sites¹⁴ (mainly serotonergic 5-HT₂ receptors; data not shown). The images of the sections of both cytoarchitectonic and autoradiographic studies were transformed into standard anatomical format by the computerized Human Brain Atlas^{14–17} so that structural findings could be compared with those of the positron emission tomography (PET) studies. An overlay map was calculated and area 4a and 4p were defined as the space of the brain within which area 4a or 4p could be found in at least three of five brains (Fig. 3a).

To test whether the two subareas differed functionally we analysed PET activations of M1 in four groups of normal right-handed subjects performing different motor actions with the fingers of the right hand (Fig. 3). The hypotheses were (1) that double representations of the fingers might exist (as suggested in refs 6, 7) and (2) that externally triggered movements and movements guided by somatosensory information might activate area 4a and 4p differentially. The PET images were transformed into the standard format of the computerized brain atlas, as were the images of the structural studies. We then checked whether the locations of the PET activations were within areas 4a or 4p (Fig. 3).

In reaction-time tasks, subjects pressed a response key with their right thumb. The somatosensory reaction-time task (response to touch on right index finger) activated two separate fields within M1, one located medially within area 4p, and one more superiorly and laterally within 4a (Fig. 3b). In this task, the movement, a thumb flexion to press the key, was not guided by the tactile stimulus, in the sense that it did not provide any information about how the thumb flexion should be performed. In a visual reaction-time task, activation of areas 4a and 4p overlapped the active regions of the somatosensory reaction-time task (Fig. 3c,f). This indicates that two different representations of the thumb might exist. In a tactile discrimination task the subjects examined ellipsoids with complex, learned, stereotypical movements in which mainly the thumb, index and middle finger were active^{18–20}. The trajectories of the fingers were guided by the curvatures of the objects^{18,19}. Although most of the activation was localized to area 4p, overlapping the activations of the reaction-time tasks (Fig. 3c), a separate overlap was also found between the 4a reaction-time task activation and the tactile discrimination task (Fig. 3f). The probability that such activations from two independent studies by chance should overlap both in area 4a and in area 4p was very small. This seems to confirm that two different representations of the thumb exist.

In a visual memory task (Fig. 3, legend), the subjects used the index and middle finger for responses. This activated area 4a in regions overlapping the tactile discrimination activation but not those associated with the reaction-time tasks (Fig. 3d,e). Finally, in a roughness discrimination task, the subjects using the thumb and index finger showed activation in area 4p in almost exactly the same territory as in the tactile discrimination task (Fig. 3c). This indicates that also two representations of the index finger and presumably also the middle finger exist: one within area 4a and one within area 4p. A clue to a possible separation of function arose from the observation that roughness discrimination did not activate area 4a, but activated area 4p significantly more than the control condition of self-generated movements ($P < 0.05$).

We have described two structurally different areas in the primary motor cortex, 4a and 4p, each having one representation of the thumb, index and possibly middle finger. These areas differ cytoarchitecturally, neurochemically, and possibly functionally, although a double dissociation of function has not been shown. □

3. Penfield, W. & Boldrey, E. *Brain* **60**, 389–443 (1937).
4. Woolsey, C. N. et al. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* **30**, 238–264 (1952).
5. Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde* (Barth, Leipzig, 1909).
6. Strick, P. L. & Preston, J. B. *J. Neurophysiol.* **48**, 139–149 (1982).
7. Strick, P. L. & Preston, J. B. *J. Neurophysiol.* **48**, 150–159 (1982).
8. Kawashima, R. et al. *Neuroreport* **6**, 238–240 (1995).
9. Matelli, M., Luppino, G., Fogassi, L. & Rizzolatti, G. *J. Comp. Neurol.* **280**, 468–488 (1989).
10. Stepniewska, I., Preuss, T. M. & Kaas, J. H. J. *Comp. Neurol.* **330**, 238–271 (1993).
11. Stepniewska, I., Preuss, T. M. & Kaas, J. H. J. *Comp. Neurol.* **349**, 558–582 (1994).
12. Stepniewska, I., Preuss, T. M. & Kaas, J. H. J. *Comp. Neurol.* **349**, 536–557 (1994).
13. Schleicher, A. & Zilles, K. *J. Microsc.* **157**, 367–381 (1990).
14. Roland, P. E. & Zilles, K. *Trends Neurosci.* **17**, 458–467 (1994).
15. Schormann, T., von Matthey, M., Dabringhaus, A. & Zilles, K. *Bioimaging* **1**, 119–128 (1993).
16. Schormann, T., Dabringhaus, A. & Zilles, K. *IEEE Trans. Med. Imag.* **14**, 25–35 (1995).
17. Roland, P. E. et al. *Hum. Brain Map.* **1**, 173–184 (1994).
18. Roland, P. E. & Mortensen, E. *Brain Res. Rev.* **12**, 1–42 (1987).
19. Seitz, R. J., Roland, P. E., Bohm, C., Greitz, T. & Stone-Elander, S. *Eur. J. Neurosci.* **3**, 481–492 (1991).
20. Kunesch, E., Binkofski, F. & Freund, H. J. *Exp. Brain Res.* **78**, 539–546 (1989).
21. Merker, B. J. *Neurosci. Methods* **9**, 235–241 (1983).
22. Mahalanobis, P. C., Majumda, D. N. & Rao, C. R. *Sankhya Kalkutta* **9**, 89–324 (1949).
23. Zilles, K. & Schleicher, A. in *Autoradiography and Correlative Imaging* (eds Stumpf, W. & Solomon, H.) 277–307 (Academic, San Diego, 1995).
24. Roland, P. E. & Larsen, B. *Arch. Neurol.* **33**, 551–558 (1976).
25. Roland, P. E., Levin, B., Kawashima, R. & Åkerman, S. *Hum. Brain Map.* **1**, 3–19 (1993).

ACKNOWLEDGEMENTS. We thank K. Rascher for help with the manuscript. This work was supported by grants from the Swedish Medical Research Council, the Deutsche Forschungsgemeinschaft and the BioTech Programme.

CORRESPONDENCE and requests for materials should be addressed to K.Z. (e-mail: zilles@hirn.uni-duesseldorf.de).

Redistribution of synaptic efficacy between neocortical pyramidal neurons

Henry Markram & Misha Tsodyks

Department of Neurobiology, The Weizmann Institute for Science, Rehovot 76100, Israel

EXPERIENCE-dependent potentiation and depression of synaptic strength has been proposed to subserve learning and memory by changing the gain of signals conveyed between neurons^{1,2}. Here we examine synaptic plasticity between individual neocortical layer-5 pyramidal neurons. We show that an increase in the synaptic response, induced by pairing action-potential activity in pre- and postsynaptic neurons, was only observed when synaptic input occurred at low frequencies. This frequency-dependent increase in synaptic responses arises because of a redistribution of the available synaptic efficacy and not because of an increase in the efficacy. Redistribution of synaptic efficacy could represent a mechanism to change the content, rather than the gain, of signals conveyed between neurons.

Changes in the amplitude of synaptic responses evoked by single-shock extracellular electrical stimulation of presynaptic fibres are usually considered to reflect a change in the gain of synaptic signals, and are the most frequently used measure for evaluating synaptic plasticity^{1–5}. Synapses, however, do not merely transmit each action potential in an identical manner but do so in a manner which is dependent on the history of activity of the synapse⁶. It is therefore not possible to extrapolate from changes in single or even paired synaptic responses to the general behaviour of the synapse. To evaluate changes in the gain of a synapse it is necessary to consider the transmission of synaptic signals more complex than single excitatory postsynaptic potentials (e.p.s.ps) such as those evoked by a train of action potentials. Test synaptic responses evoked by high-frequency trains of extracellular electrical shocks may not only alter the sensitivity of axons to subsequent stimuli, but may also recruit polysynaptic circuits. It is therefore necessary to induce high-frequency transmission

Received 20 May; accepted 1 July 1996.

1. Fritsch, G. & Hitzig, E. *Arch. Anat. Physiol. Wiss. Med.* **37**, 300–332 (1870).
2. Foerster, O. in *Handbuch der Neurologie* (eds Bumke, O. & Foerster, O.) 1–357 (Springer, Berlin, 1936).

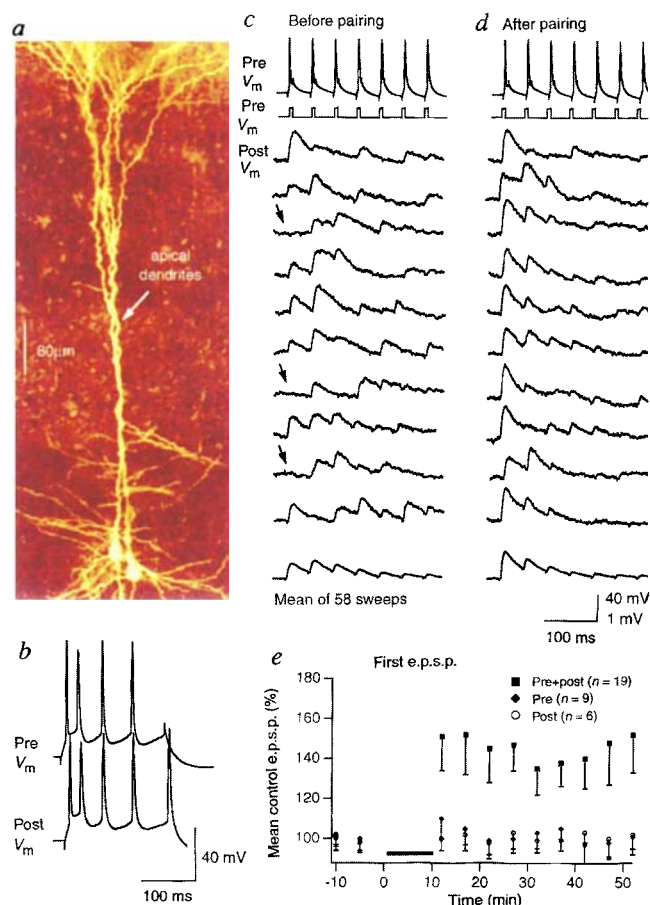
between individual neurons. A preparation was developed^{7,8} which enabled high-resolution whole-cell patch-clamp recordings of preselected layer 5 pyramidal neurons (Fig. 1a). This preparation enables repeated reliable examination of the same synaptic connection (on average 5.6 synapses per connection; H. Markram, J. Lübke, M. Frotscher, A. Roth and B. Sakmann, manuscript submitted) and has been used to demonstrate the importance of coincident synaptic input and dendritic action potentials in the induction of plasticity at this synaptic connection⁸.

A train of presynaptic action potentials at frequencies above about 0.25 Hz results in depression of the synaptic response, until a stationary level of e.p.s.p amplitude is reached (defined here as e.p.s.p_{stat}). This amplitude is reached faster and decreases as the presynaptic action-potential frequency increases. To test the effect of simultaneous burst firing of pre- and postsynaptic neurons (Fig. 1b) on subsequent transmission of trains of action potentials, 6–7 e.p.s.ps at 23 Hz were generated by brief current pulses injected into the soma of the presynaptic neuron. An example of how transmission of a high-frequency action potential train changes in this synaptic pathway after pairing pre- and postsynaptic firing, is shown in Fig. 1. During the control period, the synapse seemed to respond randomly to the train (Fig. 1c). After pairing, the first action potential in the train almost always generated the largest response (Fig. 1d). The increase in this initial e.p.s.p (e.p.s.p_{init}) lasted more than 40–60 min (Fig. 1e) consistent with a recent study of these synapses⁸. The effect of pairing was hebbian in the sense that burst firing in only the presynaptic (9 synaptic connections) or in only the postsynaptic (5

synaptic connections) neuron failed to induce comparable changes (Fig. 1e). The increase in the single action-potential-evoked e.p.s.p induced by the same pairing protocol is also prevented when *N*-methyl-D-aspartate (NMDA) receptors are blocked at these synapses⁸. On the basis of the changes in e.p.s.p_{init} alone it appears that the gain of the synaptic connection might have increased. The entire synaptic response to the action potential train, however, was not uniformly increased, indicating a more complex effect on transmission.

A further examination of the change induced by pairing revealed that whereas the amplitude of e.p.s.p_{init} was increased in 18 of 19 synaptic connections ($66.5 \pm 17.08\%$; $n = 19$; mean $\pm$ s.e.m.; $P < 0.001$) the average amplitude of e.p.s.p_{stat} (see Fig. 2b) was unaffected (Fig. 2a–c) indicating that the absolute efficacy of the synaptic connection was unaffected. The effect of pairing on the amplitude of the e.p.s.ps generated before the stationary level was reached (defined here as transition e.p.s.ps; see Fig. 2b) and consequently the area below the voltage response for the entire train was either increased, decreased or unaffected (no mean change; range $< -50\%$ to $> +50\%$ change). The effect of pairing on e.p.s.p_{init} and the lack of effect on e.p.s.p_{stat} was also observed when extracellular Ca^{2+} concentration was lowered to the physiological level for a rat of this age (1.5 mM (ref. 9); 6 synaptic connections; Fig. 2c). The higher apparent failures to evoke an e.p.s.p at lowered extracellular $[\text{Ca}^{2+}]$ also enabled detection of a pairing-induced decrease in failure rate (from $21 \pm 5.4\%$ to $5 \pm 1.8\%$) which was not significant at higher $[\text{Ca}^{2+}]_o$ because of very few initial apparent failures ($2.6 \pm 2.17\%$

FIG. 1 The effect of pairing pre- and postsynaptic burst firing on high-frequency transmission. a, The cell type. A pseudo-coloured collage image of a pair of thick-tufted layer 5 pyramidal neurons loaded with biocytin illustrates the cell type from which all paired recordings were made. b, The pairing method. Typically, 200-ms sustained current pulses were co-injected into pre- and postsynaptic neurons with sufficient current to evoke 4–8 spikes. Current pulse in the postsynaptic neurons was delayed (1–5 ms) to ensure that the postsynaptic neuron discharged after onset of synaptic input. No attempt was made to control subsequent spikes. The procedure was repeated 30 times every 20 s. c, Test-train responses. A precisely controlled train of presynaptic spikes (23 Hz) (top voltage trace) was generated every 5 s by injecting 2 nA, 5 ms current pulses into the soma of the presynaptic neuron. Responses to action potential trains (postsynaptic membrane voltage (Post V_m) traces). Arrows mark failures of first spike to evoke an e.p.s.p. Control failure rate for the first e.p.s.p. was 12.8% as determined from 58 consecutive sweeps (mean shown in lower trace). d, Train response after pairing. The first spike in the train almost always evoked the largest e.p.s.p. Failure rate decreased to 1.7%. e, Effect of pairing is lasting and hebbian. Changes in mean amplitude (see Methods) of e.p.s.p_{init} at various times before and after pairing. Separate populations of synaptic connections tested for effects of pairing (Pre + Post), presynaptic burst firing alone (Pre) and postsynaptic burst firing alone (Post). Bar indicates the pairing period. METHODS. Sagittal slices (300 μm) were cut from the neocortex of Wistar rats (13–15 d) as described in H. Markram *et al.* (manuscript submitted). All experiments were done at 30–32 °C within 6.5 h after slicing. The extracellular solution contained (in mM): 125 NaCl, 2.5 KCl, 25 glucose, 25 NaHCO_3 , 1.25 NaH_2PO_4 , 2 CaCl_2 and 1 MgCl_2 . Layer 5 pyramidal neurons from the somatosensory cortical area were identified using infrared differential interference contrast video-microscopy on an upright microscope (Zeiss-Axiokop-FS, fitted with $\times 40$ -W/0.75NA objective) as described¹⁶. Somatic whole-cell recordings (10–20 M Ω access resistance) signals were amplified using two Axoclamp-2B amplifiers (Axon Instruments) and captured on computer using pulse control (by R. Bookman and co-workers, Miami University) and analysed using IGBOR (Igor Wavemetrics, Lake Oswego, OR). Recordings were made using pipettes containing (in mM): 100K-gluconate, 20 KCl, 4 ATP-Mg, 10 phosphocreatine, 0.03 GTP, 10 HEPES buffer and 0.5% biocytin (pH 7.3, 310 mOsm). Resting membrane-potential levels typically -62 ± 2 mV. Independent and paired *t*-tests were used to determine significance of changes. All data are presented as mean $\pm$ s.e.m. Data after pairing or after control protocol represent maximal changes (positive or negative) recorded over a 5-min period (20–60 sweeps) 10–40 min after the end of the pairing period. Whole-cell recordings from two cells were made virtually simultaneously after the configuration was established on both neurons. Pairing was initiated 10–15 min after whole-cell recording. Input and access resistances were monitored by producing brief hyperpolarizing or



depolarizing current pulses in the somata after every synaptic response. Responses to presynaptic spike trains were recorded every 15 s in 55 of 59 synaptic connections examined and every 5 s in the remaining 4. Eight synaptic connections were discarded because of unreliable mean sweeps caused by excessively high variation or unstable membrane potentials.

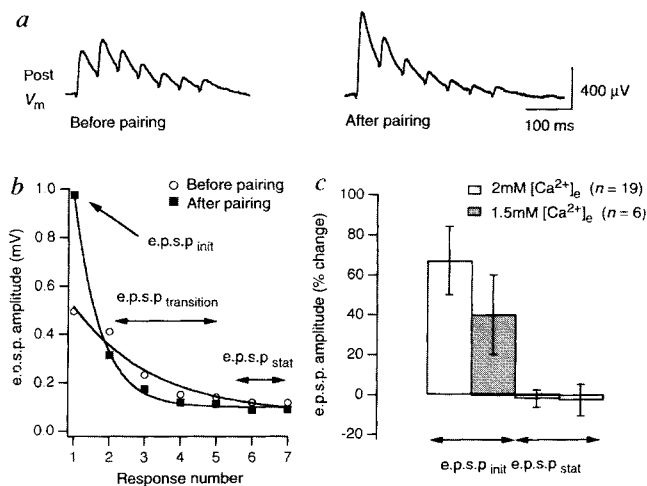


FIG. 2 Pairing induces an increase in initial e.p.s.ps and has no effect on stationary e.p.s.ps. *a*, The mean of 58 sweeps during the control period (~ 5 min; left trace) and maximally changed mean of 59 sweeps recorded 20 min after pairing. *b*, e.p.s.p. amplitudes in *a* were measured from the voltage immediately before the onset of the e.p.s.p. to the peak of the e.p.s.p. Superimposed is a single exponential fit. Defined are initial, transition and stationary e.p.s.ps. The amplitude of e.p.s.p. p_{stat} was taken as the average of the last 50 points of the fitted curve which was equivalent to the last 20% of the curve and roughly equivalent to the average of the last 2 e.p.s.ps. This approach was adopted because the coefficient of variation of later e.p.s.ps is typically above 100% (compared to around 40% for e.p.s.p. p_{init}) and the amplitudes of individual e.p.s.ps are therefore less reliable given the restricted number of sweeps. The accuracy of this approach to reflect the amplitude of e.p.s.p. p_{stat} was empirically confirmed by averaging progressively larger number of sweeps and comparing the amplitudes. *c*, e.p.s.p. p_{init} and e.p.s.p. p_{stat} in two separate synaptic populations in which $[Ca^{2+}]_o$ was either 2 or 1.5 mM. In 2 mM $[Ca^{2+}]_o$ pre-pairing mean of e.p.s.p. p_{init} is 1.17 ± 0.229 mV and e.p.s.p. p_{stat} is 184 ± 28 μ V; $n = 19$.

before pairing and $0.33 \pm 0.25\%$ after pairing).

A more rapid depression of synaptic responses during the spike train is characteristic of the change induced by pairing (see Fig. 2b). The time constant for the decay of the amplitudes (defined as DTC) of successive e.p.s.ps decreased by 30% from (52 ± 6.7 ms to 36 ± 4.3 ms) whereas the time constant for e.p.s.p. p_{init} to recover to its maximum was unaffected by pairing (1.135 ± 0.034 ms before and 1.133 ± 0.051 ms after; 11 synaptic connections). The DTC value before pairing indicated the status of the synaptic connection and could therefore be used to predict quite accurately the change in e.p.s.p. p_{init} and in the DTC after pairing (positive correlation between pre-pairing DTC and change in e.p.s.p. p_{init} ; $r = 0.92$). The natural range of DTC values observed before pairing is most likely due to differences in the use of the available efficacy by each synaptic connection (leaving more or less efficacy for subsequent action potentials) which could result from a continuum of release probabilities within a synaptic population¹⁰.

The increase in the rate of synaptic depression after pairing was not caused by indirect effects on polysynaptic transmission as synaptic depression was not affected by the blockade of γ -amino butyric acid (GABA) receptors with picrotoxin (5 μ M, Sigma) and saclophen (10 μ M; Tocris) (3 synaptic connections; not shown) or by blockade of metabotropic glutamate receptors with 2-amino-5-phosphonopropionic acid (10 μ M; Tocris; 2 synaptic connections; not shown). Synaptic depression was also not affected when postsynaptic voltage-activated channels were blocked after repatching and loading neurons with the lidocaine derivative, *N*-ethylbromide quaternary salt (QX-314, 5 μ M, Research Biochemical Inc., 5 synaptic connections, not shown).

The increase in the amplitude of e.p.s.p. p_{init} , which actually represents e.p.s.p. p_{stat} for very low-frequency stimulation (< 0.25 Hz),

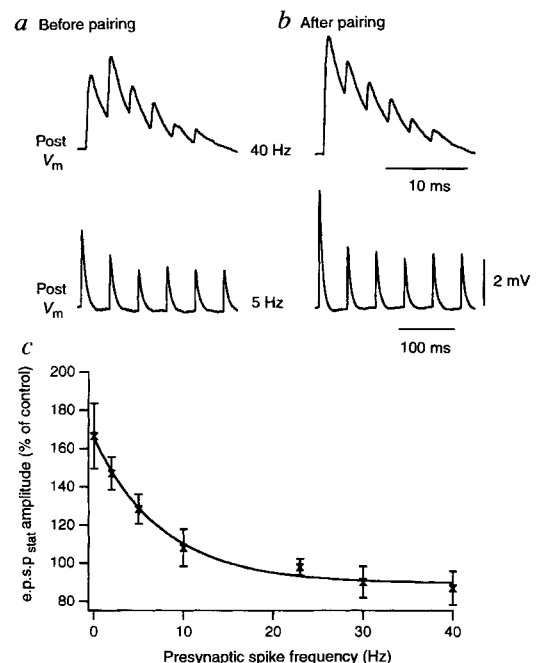


FIG. 3 Frequency dependence of synaptic potentiation. *a*, *b*, Results from an experiment on a single synaptic connection in which e.p.s.ps were generated alternately by 5-Hz and 40-Hz presynaptic spike trains before and after pairing. e.p.s.p. p_{init} is increased to the same extent for both frequencies, e.p.s.p. p_{stat} is not changed in the 40-Hz trace but is increased in the 5-Hz trace (compare for example the amplitude of the last 2 e.p.s.ps in the 40-Hz train before and after pairing). Even at 5 Hz, the e.p.s.p. p_{stat} was only increased by 25% whereas e.p.s.p. p_{init} was increased by 58%. *c*, Summary of the change in e.p.s.p. p_{stat} . Total of 33 synaptic connections represented (1–4 frequencies tested per synaptic connection). The first, leftmost point represents very-low-frequency presynaptic spikes (0.25 Hz in 2 synaptic connections and 0.067 Hz in 17 synaptic connections) and actually are the amplitudes of e.p.s.p. p_{init} of Fig. 2. Second point, 2 Hz ($n = 11$). Third point, 5 Hz ($n = 10$). Fourth point, 23 Hz ($n = 19$). Fifth point 30 Hz ($n = 4$) and last point, 40 Hz ($n = 9$).

and the lack of an effect on the amplitude of e.p.s.p. p_{stat} for a high-frequency train, indicates that the potentiation is conditional on the presynaptic spike frequency. The effect of pairing on e.p.s.p. p_{stat} at several different frequencies was therefore examined. In the same neuron, e.p.s.p. p_{init} was found to increase equally for low- and high-frequency trains, whereas e.p.s.p. p_{stat} was increased only when the frequency was low (Fig. 3a, b). Potentiation of synaptic responses therefore only occurred when the presynaptic frequency was below 20 Hz (Fig. 3c).

The lack of an effect on the e.p.s.p. p_{stat} level indicates that postsynaptic receptors were not recruited or potentiated^{11–13} nor were silent synapses unmasked¹⁴. As pairing did not affect the rate of recovery from depression, it would seem that the actual biophysical mechanism of synaptic depression was also unaffected by pairing and that the increased rate of depression was secondary to increased use of the available efficacy of the synaptic connection by action potentials. The most likely mechanism for increased use of the existing synaptic efficacy is an increase in the probability of transmitter release after pairing (see also ref. 15), but it could also be generated by an increase in the affinity of postsynaptic receptors for glutamate if these receptors are not saturated.

The effect reported here is not an unconditional potentiation of the efficacy of the synaptic connection; instead, it is a redistribution of the existing efficacy between spikes in a train. This type of plasticity should be distinguished from unqualified potentiation or depression of synaptic efficacy, which imply a change in the gain of synaptic signals transmitted and which would be produced by

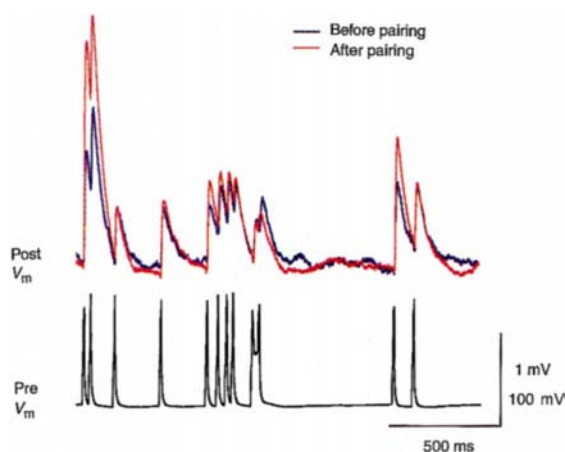


FIG. 4 Complex effect of pairing on the synaptic response generated by an irregular presynaptic action potential train. An irregular spike train was generated by brief (5 ms) current pulses repeatedly for 30 episodes and the postsynaptic response averaged before (blue) and after (red) pairing.

mechanisms such as postsynaptic changes or by adding or removing synapses to or from the connection. Redistribution of synaptic efficacy could potentially occur at any synapse in the brain if activated at a faster rate than required for complete recovery of the synaptic efficacy.

The physiological implications of redistribution of efficacy are also entirely different from unconditional potentiation or depression and are partly predicted by the frequency-dependent potentiation seen in Fig. 3. Under *in vivo* conditions, neurons tend to discharge irregularly, which effectively represents a multitude of spike frequencies persisting for different time periods, indicating that the effect of pairing on synaptic input generated by an irregular presynaptic spike train would be complex if redistribution of synaptic efficacy was to occur (Fig. 4). The effect cannot be predicted as most synaptic responses during such a train have not reached a stationary level for the given frequency and hence are all transition e.p.s.ps (see Fig. 2b). As discussed, transition e.p.s.ps could be enhanced, depressed or unchanged after pairing. Redistribution of synaptic efficacy may therefore serve as a powerful mechanism to alter the dynamics of synaptic transmission in subtle ways and hence to alter the content rather than the gain of signals conveyed between neurons. □

A diffusible coupling signal from the transplanted suprachiasmatic nucleus controlling circadian locomotor rhythms

Rae Silver, Joseph LeSauter, Patrick A. Tresco* & Michael N. Lehman†

Departments of Psychology, Barnard College and Columbia University, 3009 Broadway, New York, New York 10027, USA

* Department of Bioengineering, Room 506 BPR, University of Utah, Salt Lake City, Utah 84112, USA

† Department of Cell Biology, Neurobiology and Anatomy, University of Cincinnati, College of Medicine, Cincinnati, Ohio 45267, USA

THE mammalian suprachiasmatic nuclei (SCN) transmit signals to the rest of the brain, organizing circadian rhythms throughout the body¹⁻⁴. Transplants of the SCN restore circadian activity rhythms to animals whose own SCN have been ablated⁵⁻⁹. The nature of the coupling signal from the grafted SCN to the host brain is not known, although it has been presumed that functional recovery requires re-establishment of appropriate synaptic connections. We have isolated SCN tissue from hamsters within a semipermeable polymeric capsule before transplantation, thereby preventing neural outgrowth but allowing diffusion of humoral signals. Here we show that the transplanted SCN, like neural pacemakers of *Drosophila*¹⁰ and silkworms¹¹, can sustain circadian activity rhythms by means of a diffusible signal.

To show that the recovered circadian rhythm is produced by the donor tissue and is not a consequence of spontaneous recovery of function, the host animals used were male hamsters bearing the *tau* mutation¹² and were raised in our colony. This mutation has the behavioural effect of altering the period of the endogenous circadian rhythm to about 22 hours in heterozygotes and 20 hours in homozygotes. Donor animals were fetuses (embryonic day 14–15) from wild-type, time-pregnant females (Charles River, MA). Because of the difference in period length between host and donor phenotype, the period of the locomotor rhythm could be unambiguously attributed to either the donor SCN or the host animal¹³.

Animals were housed individually in cages equipped with running wheels, and were allowed unlimited access to food and water. The cages were placed in a room in constant darkness (DD) equipped with a dim red light (<1 lux; Delta 1, Dallas) to allow for animal maintenance and a white-noise generator (91 dB) to mask environmental noise, and maintained at about 23 °C. After the host rhythm had been characterized (for 7–14 days), hamsters were anaesthetized with pentobarbital (100 mg per kg) and bilateral lesions of the SCN were made as previously described⁹. Animals in which 24-hour periodicity was not seen upon inspection of the actogram were used as hosts for implantation (see Fig. 1 legend). The extent of the lesion was determined histologically at the end of the experiment.

After behavioural testing, hamsters were heavily anaesthetized (pentobarbital 200 mg per kg) and perfused intracardially with saline followed by 4% paraformaldehyde in 0.1 M phosphate buffer, pH 7.3. Coronal sections (20 µm) were cut on a cryostat and alternate sections were immunostained for vasoactive intestinal polypeptide (VIP), vasopressin-associated neurophysin (NP), cholecystokinin (CCK), or calbindin-D28K protein (CaBP), to verify the lesion and the presence of neurochemicals characteristic of the SCN in the encapsulated graft. Polyclonal antisera to VIP, NP, CCK (Incstar, MN) and CaBP (Sigma, MO) were obtained and these antigens were detected using a modified avidin-biotin-HRP procedure (Vectastain Elite Kit, Vector Laboratories, CA).

Received 21 February; accepted 28 June 1996.

1. Teyler, T. J. & Discenna, P. *Brain Res.* **319**, 15–28 (1984).
2. Barnes, C. A. *Neuron* **15**, 751–754 (1995).
3. Stevens, C. F. *Neuron* **10**, 55–63 (1993).
4. Kullman, D. M. & Siegelbaum, S. A. *Neuron* **15**, 997–1002 (1995).
5. Bliss, T. V. P. & Lomo, T. J. *Physiol. (Lond.)* **232**, 331–336 (1973).
6. Thomson, A. M. & Deuchars, J. *Trends Neurosci.* **17**, 119–126 (1994).
7. Markram, H., Lübke, J., Sakmann, B. & Frotscher, M. in *European Neuroscience Association Abstr.* **34.27** (1995).
8. Markram, H. & Sakmann, B. in *Soc. Neurosci. Abstr.* **788.11** (1995).
9. Jones, H. C. & Keep, R. F. *J. Physiol. (Lond.)* **P64** (1987).
10. Allen, C. & Stevens, C. F. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10380–10383 (1994).
11. McNaughton, B. L. *J. Physiol. (Lond.)* **324**, 249–262 (1982).
12. Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Neuron* **1**, 911–917 (1988).
13. Bashir, Z. I., Alford, S., Davies, S. N., Randall, A. D. & Collingridge, G. L. *Nature* **349**, 156–158 (1991).
14. Liao, D. Z., Hessler, N. A. & Malinow, R. *Nature* **375**, 400–404 (1995).
15. Stevens, C. F. & Wang, Y. *Nature* **371**, 704–707 (1994).
16. Stuart, G. J., Dodt, H.-U. & Sakmann, B. *Pflügers Archiv.* **423**, 511–518 (1992).

ACKNOWLEDGEMENTS. We thank G. Borst for helpful comments on the study; M. Segal, E. Ahissar, A. Aertsen, Y. Dudai, R. Rahamimoff, W. Singer, B. Katz, R. Douglas, and I. Segev for their comments on the manuscript; D. Pikus for programming and data analysis and A. Gupta for histology. H.M. holds the Joseph D. Shane career-development chair. The study was supported by research grants from the Henry S. and Ann S. Reich Research Fund for Mental Health, the Minerva Foundation and the ONR.

CORRESPONDENCE and requests for materials should be addressed to H.M. (e-mail: bnmark@weizmann.weizmann.ac.il).

Horseradish peroxidase (HRP) was demonstrated using diaminobenzidine (DAB) as the chromogen. Some slides were counterstained using cresyl violet (Chroma-Gesellschaft, Germany).

Figure 1 shows a representative animal with a complete SCN lesion in which locomotor activity was restored. The endogenous rhythm of the mutant host is 21.4 hours, and the animal becomes arrhythmic after ablation of the SCN. Roughly 21 days after transplantation of the encapsulated graft, the locomotor rhythm recorded was that of the donor, with a period of 24.0 hours. It has previously been shown that in animals with partial damage to their own SCN, implantation of grafts taken from donors with a different circadian period results in 'temporal chimaeras'¹⁴. In such animals, the circadian period of the donor and the host are each expressed in the activity record. Figure 2 depicts a temporal

chimaera animal bearing a partial SCN lesion. The host rhythm was 21.9 hours, and after transplantation of the encapsulated graft, two distinct free-running rhythms were simultaneously expressed, with periods of about 24.0 (donor) and 21.8 (host). The data for all the animals (summarized in Fig. 3), show that using wild-type donor tissue, the period of the restored rhythm is roughly 24 hours (solid line), whether the host animal is heterozygous ($n = 7$) or homozygous ($n = 1$) for the *tau* mutation. Of the eight recovered animals, six had complete destruction of the host SCN as indicated by the absence of all peptidergic markers listed above in cells or fibres at the site of the lesion. In these animals, the donor, but not the host period is observed. In two animals with partial damage to the SCN, the period of both host (dashed line) and donor (solid line) locomotor rhythm is seen.

One capsule was lost before processing. The remaining seven capsules each contained viable Nissl-stained tissue and VIP-, NP-, CCK- and CaBP-positive fibres and/or cells (Fig. 4). The capsules were located either in the third ventricle at the level of the anterior hypothalamus, or in the Foramen of Monro. Neurochemically identified cells and fibres were seen in a characteristic organization resembling the intact hamster SCN. For example, CaBP-positive cells and fibres were found in non-overlapping portions of the graft to CCK fibres (Fig. 4c, d). Six of seven capsules were completely closed (for example, Fig. 4a, b). In the remaining animal, one end of the capsule had a small aperture (about 40–70 μ m diameter) but the tissue was at the distal end of the capsule, several millimeters away from the opening. There was no connection between graft and host in any of these animals. Restored function was not seen in any animal lacking viable tissue in the capsule. In seven animals, tissue was present in the capsule, but no restoration of function was seen. In five of these, the encapsulated

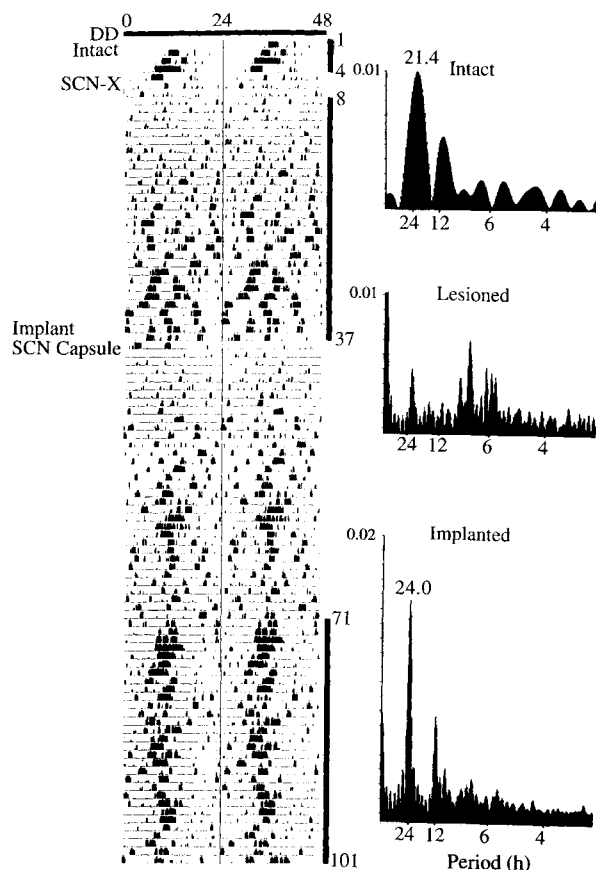


FIG. 1 Expression of wild-type rhythmicity in a grafted, mutant hamster with total ablation of the SCN. The endogenous rhythm of the intact mutant host is seen at the top of the record (period 21.4 h). After SCN ablation (SCN-X), the animal becomes arrhythmic. Implantation of the encapsulated graft occurs on day 37 of the record, and gradual emergence of circadian rhythmicity with the donor period (24 h) is seen in the following weeks. The black vertical bars on the right of the actogram indicate the days (intact, lesioned and implanted) on which spectral analyses were done.

METHODS. Transplants were done 2–5 weeks after SCN ablation. The pregnant hamster was anaesthetized intracardially with 200 mg per kg pentobarbital. Pups were rapidly removed, and donor SCN were micro-dissected (under aseptic conditions) from the ventral surface of the hypothalamus and placed in sterile saline in a sterile petri dish on ice until the dissection of the entire litter was complete. A length of polyethylene tubing (5 cm long; outer diameter 0.97 mm) attached to a Hamilton syringe was inserted in one end of the hollow fibre and the tissue (in a medium of 50% matrigel (Gibco BRL) and 50% alginate (2% in phosphate buffered saline; Kelcogel HV, Kelco, NJ) was drawn up by suction. The fibre was then cut every 5 mm, heat-sealed and implanted stereotactically in the third ventricle of SCN-lesioned hamsters. After recovery, hamsters were returned to their cages and wheel running behaviour was monitored for 8–16 weeks.

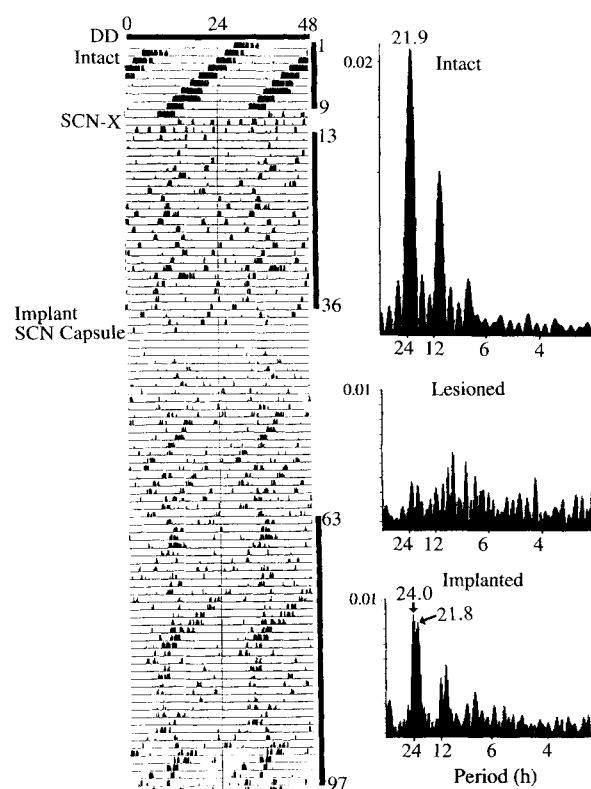


FIG. 2 Expression of wild-type rhythmicity in a grafted, mutant hamster with partial ablation of the SCN. After lesion placement (SCN-X, day 10) the animal became arrhythmic. By about day 63 of the record, locomotor rhythmicity with the periods of both the donor (24 h) and the host (21.8 h) can be seen. Spectral analyses of the data are shown on the right of the actogram, as in Fig. 1.

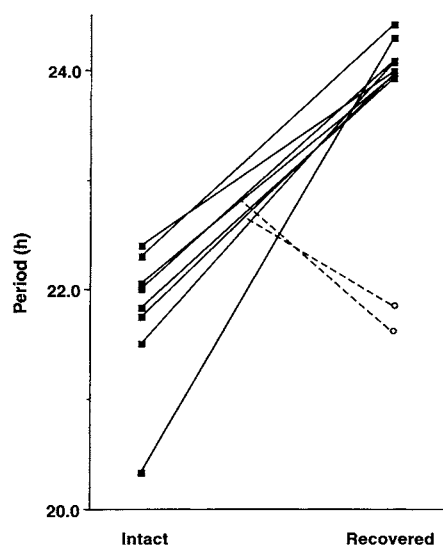


FIG. 3 Transplantation of encapsulated donor graft results in expression of the donor period. For six of the host animals, the endogenous rhythm (left) was eliminated and the donor rhythm (right) was restored by the graft. In two animals with partial SCN lesions (dashed lines), a chimaeric rhythm was expressed after grafting.

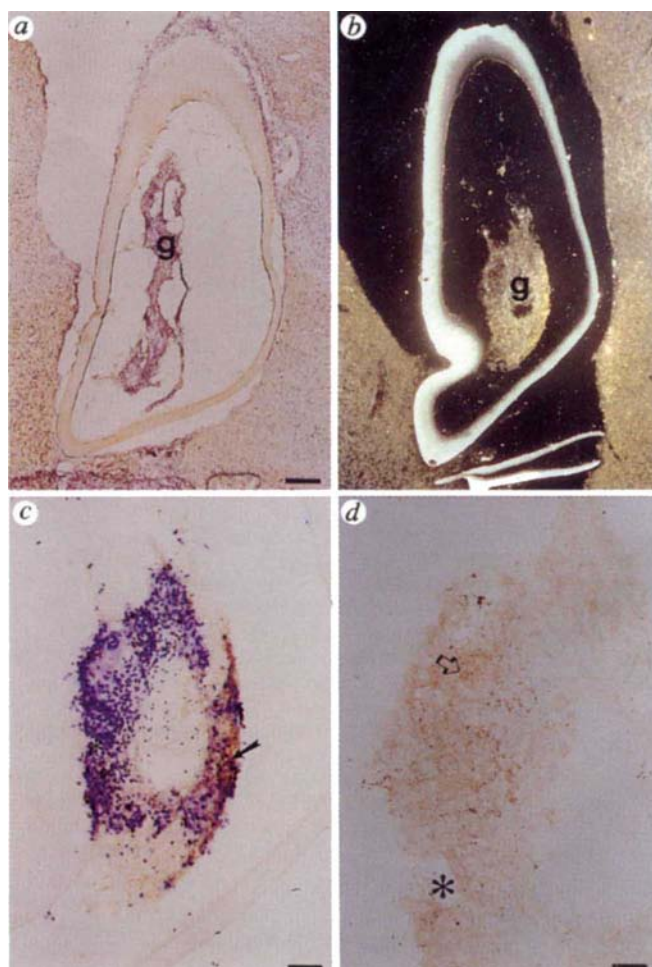
METHODS. Computer-based systems were used to monitor locomotor activity for power-spectrum analysis (Dataquest III, Data Sciences Inc., St. Paul, MN) and for the calculation of the precise period of the locomotor rhythm (tau program, Mini Mitter, Sunriver, OR). After the SCN lesions, some hamsters exhibited transient ultradian or circadian components lasting for a few cycles, as previously reported²². In contrast, the recovered circadian rhythm was sustained, and lasted for the duration of the study. The period of the recovered rhythms ranged from 23.93 to 24.41 h. The time of onset of locomotor activity differed among animals, and was distributed throughout the 24 h cycle, reflecting the absence of entraining signals.

tissue was mostly either necrotic or gliotic, and no VIP-, NP-, CCK- and CaBP-positive neurons or axons were found. In two non-recovered animals, tissue similar to that of recovered animals was seen.

Previous work indicated that circadian rhythms of multiunit electrical activity survived within a 'hypothalamic island' containing the SCN but were lost in each of the areas sampled outside of

the circular knife cut¹⁵. The results were interpreted as indicating that neural efferents are necessary for circadian rhythmicity. After similar isolation from the rest of the brain, the isolated SCN can, however, sustain circadian locomotor and endocrine rhythms^{16,17}. Also, SCN transplants with varied attachment sites and efferent connections to the adjacent host tissue in the third ventricle effectively restore locomotor, drinking, and gnawing rhythms^{6,9,18}. Our finding shows how circadian rhythmicity of the SCN can be relayed to targets. The results demonstrate that a diffusible signal is sufficient to restore rhythmicity. It may be that a diffusible signal normally travels across synapses, but in the isolated SCN it reaches targets via the CSF or the extracellular space. Alternatively, in the intact animal, a specific rhythm may depend on neural efferent or

FIG. 4 Anatomy of encapsulated SCN grafts. *a, b*, Low-power photomicrographs of encapsulated grafts (g). The actogram from the animal in *a* is shown in Fig. 2. The sections have been immunostained for CaBP and counterstained with cresyl violet, and are viewed under either bright-field (*a*) or dark-field illumination (*b*). CaBP-positive fibres and cells, which appear yellow-gold under dark-field illumination, are seen in both the graft and the host brain. Scale bar, 200 μ m. *c*, Bright-field illumination of the same graft shown in *b*. This doughnut-shaped graft consists of areas of Nissl-stained cells and CaBP-positive fibres and neurons (example indicated by arrow), as well as areas where Nissl or CaBP-stained cells are sparse. Scale bar, 80 μ m. *d*, Higher-power view of the adjacent section to that shown in *c*, immunostained for CCK. Immunopositive fibres and a few lightly stained cell bodies (example indicated by arrow) are seen in the graft. Note that the area where CCK fibres and cells are found contains few CaBP-positive fibres (compare with *b* and *c*), resembling the neurochemical organization seen in the intact hamster SCN. For orientation, asterisks in *c* and *d* indicate an indentation of the graft surface seen in both sections. Scale bar, 40 μ m. **METHODS.** For capsule fabrication, polymer solutions consisted of either 15% poly(acrylonitrile-vinylchloride) (PAN-PVC) in dimethylformide(DMF) (W/V) or a blended solution of 11% PAN-PVC and 4% polyethylene oxide (PEO) in DMF (W/V) held at room temperature. The PAN-PVC/PEO material was used in 5 of the 8 animals with functional grafts. Semipermeable tubular encapsulating membranes with an internal diameter of $\sim$ 800 μ m and a wall thickness of 65 μ m were fabricated by a co-extrusion, dry-jet wet spinning-phase inversion method as previously described^{23,24}. Scanning atomic force microscopy of both membranes indicated that rejection-layer pore size ranged from 5–120 nm. Convective and diffusional analysis of membrane permeability indicated that both membranes were of the ultrafiltration-type. The 15% PAN-PVC membrane excluded molecules with a mass greater than 100K, whereas the 11% PAN-PVC membrane and 4% PEO membrane rejected solutes above 500K.



diffusible signals or both. Some functions (not restored by neural transplants), such as reproductive responses to daylength⁹ and endocrine rhythms¹⁹ may require neural efferents. Thus the pineal melatonin response to day length depends on a known neural pathway from the SCN to the pineal gland, and is not restored by SCN transplants. The presence of a diffusible coupling signal implies that resetting of disrupted activity–rest cycles, such as those of long distance travellers ('jet-lag') and shift workers, can be achieved by a direct action of an as yet unidentified neurochemical acting directly on neural targets. Such an agent (or agents) may also influence sleep disturbances seen in naturally occurring states such as ageing²⁰, and in pathological conditions including manic depressive disorders²¹. □

Received 16 May; accepted 27 June 1996.

1. Klein, D. C., Moore, R. Y. & Reppert, S. M. *The Suprachiasmatic Nucleus: The Mind's Clock* (Oxford Univ. Press, New York, 1991).
2. Morin, L. P. *Brain Res. Rev.* **67**, 102–127 (1994).
3. Stephan, F. K. & Zucker, I. *Proc. Natl Acad. Sci. U.S.A.* **69**, 1583–1586 (1972).
4. Mosko, S. S. & Moore, R. Y. *Proc. Natl Acad. Sci. U.S.A.* **75**, 6243–6246 (1978).
5. Silver, R., Lehman, M. N., Gibson, M., Gladstone, W. R. & Bittman, E. L. *Brain Res.* **525**, 45–58 (1990).

6. Aguilar-Roblero, R. A., Morin, L. P. & Moore, R. Y. *Exp. Neurol.* **130**, 250–260 (1994).
7. DeCoursey, P. J. & Buggy, J. *Brain Res.* **500**, 263–275 (1989).
8. Sollars, P. J., Kimble, D. P. & Pickard, G. E. *J. Neurosci.* **15**, 2109–2122 (1995).
9. Lehman, M. N., Silver, R., Gladstone, W. R., Kahn, R. M., Gibson, M. & Bittman, E. L. *J. Neurosci.* **7**, 1626–1638 (1987).
10. Handler, A. M. & Konopka, R. J. *Nature* **279**, 236–238 (1979).
11. Truman, J. W. & Riddiford, L. M. *Science* **167**, 1624–1626 (1970).
12. Ralph, M. R. & Menaker, M. *Science* **241**, 1225–1227 (1988).
13. Ralph, M. R., Foster, R. G., Davis, F. C. & Menaker, M. *Science* **247**, 975–978 (1990).
14. Vogelbaum, M. A. & Menaker, M. *J. Neurosci.* **12**, 3619–3627 (1992).
15. Inouye, S. T. & Kawamura, H. *Proc. Natl Acad. Sci. U.S.A.* **76**, 5962–5966 (1979).
16. Hakim, H., DeBernardo, A. P. & Silver, R. J. *Biol. Rhythms* **6**, 97–113 (1991).
17. Honma, K., Honma, S. & Hiroshige, T. *Am. J. Physiol.* **246**, 949–954 (1984).
18. LeSauter, J. & Silver, R. *Restor. Neurol. Neurosci.* **6**, 135–143 (1994).
19. Bittman, E. L., Matsumoto, S., Markuns, J., Meyer, E. & Jetton, E. E. *Soc. Neurosci. Abstr.* **19**, 236.17 (1993).
20. Turek, F. W., Penev, P., Zhang, Y., van Reeth, O. & Zee, P. *Neurosci. Biobehav. Rev.* **19**, 53–58 (1995).
21. Mahé, V. & Chevalier, J. F. *Presse Med.* **24**, 1041–1046 (1995).
22. Rusak, B. J. *Comp. Physiol.* **118**, 145–164 (1977).
23. Aebischer, P., Wahlberg, P., Tresco, P. & Winn, S. R. *Biomaterials* **12**, 50–56 (1991).
24. Tresco, P. A. et al. *Cell Transplant.* **1**, 255–264 (1992).

ACKNOWLEDGEMENTS. The research was supported by grants from the NIH to R.S. and M.N.L. and AFOSR to R.S. and P.T.

CORRESPONDENCE and requests for materials should be addressed to R.S. (e-mail: QR@columbia.edu).

Activation of Raf by ionizing radiation

Usha Kasid*†, Simeng Suy*, Paul Dent‡§, Sugata Ray*, Theresa L. Whiteside|| & Thomas W. Sturgill‡

Departments of * Radiation Medicine, and † Biochemistry and Molecular Biology, Lombardi Cancer Center, Georgetown University, Washington DC 20007, USA

‡ Howard Hughes Medical Institute and Markey Center for Cell Signalling, Departments of Medicine and Pharmacology, University of Virginia, Charlottesville, Virginia 22908, USA

|| Department of Pathology, The Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, Philadelphia 15261, USA

THE critical pathways through which ionizing radiation induces malignant transformation and cell death are not well defined. Raf-1, a cytoplasmic serine–threonine protein kinase, mediates the transmission of mitogenic signals initiated at the cell membrane to the nucleus, resulting in the activation of transcription factors that regulate cell growth and proliferation^{1,2}. Moreover, Raf-1 overexpression and activation increases the survival response of mammalian cells to the toxic effects of ionizing radiation by an as-yet unknown mechanism (refs 3, 4 and V. Soldatenkov et al.; manuscript submitted). Somewhat analogous to mitogen-induced signalling, radiation stimulates protein-tyrosine kinase(s) and transcription factors^{5,6}. No direct biochemical link has been established, however, between radiation-stimulated protein tyrosine phosphorylation and downstream signals. Here we report a series of radiation-responsive events in which protein-tyrosine phosphorylation is followed by membrane recruitment, then tyrosine phosphorylation and activation of Raf-1 *in vivo*. Our results show that radiation-stimulated protein-tyrosine kinase(s) modify Raf-1, and implicate Raf-1 in the ionizing-radiation signal-transduction pathway.

The exposure of human laryngeal squamous carcinoma cells, PCI-04A, to ionizing radiation (15 Gy, 114 cGy min⁻¹) resulted in the temporal induction of tyrosine phosphorylation of Raf-1 protein serine–threonine kinase (Fig. 1a, b). The increase in tyrosine phosphorylation of Raf-1 quantified by densitometry was two- to threefold (Fig. 1b). Western blotting confirmed that the tyrosine-phosphorylated 75K protein immunoprecipitated by

anti-phosphotyrosine (pTyr) antibody was recognized by anti-Raf-1 antibody (data not shown). Furthermore, no changes in the total amounts of Raf-1 protein *per se* were detected after irradiation (Fig. 1c).

Tyrosine phosphorylation or acidic substitution of Raf-1 tyrosine residues at positions 340 and 341 (Y340/341) stimulates Raf-1 enzymatic activity⁷. To determine whether irradiation-stimulated phosphorylation of Raf-1 resulted in its catalytic activation, we measured Raf-1 activity in immune complexes with mitogen-activated protein-kinase kinase-1 (MAP-kinase kinase-1, MKK1) as substrate, and by coupled MKK-dependent phosphorylation of a mutant 42K MAP kinase, in which a lysine residue at position 52 is replaced by asparagine (K52R p42^{mapk}). Consistent with induction of Raf-1-tyrosine phosphorylation, Raf-1 protein-kinase activity increased after irradiation (Fig. 1d, e). We also observed an increase in the *in vitro* autokinase activity of Raf-1 in irradiated cells, although this was less than MKK1 phosphorylation (Fig. 1d). In addition, irradiation stimulated the activities of MKK and MAP kinase, which function downstream of Raf-1 in this signalling pathway (Fig. 1e). To confirm the importance of irradiation-stimulated tyrosine phosphorylation in Raf-1 activation, we tested the ability of a recombinant, truncated protein-tyrosine phosphatase 1B (PTP-1B) to reverse the stimulation of Raf-1 activity resulting from irradiation. Treatment of immunoprecipitated Raf-1 with PTP-1B decreased its activity by 50–80% (Fig. 1f). Raf-1 inactivation was attributable to PTP-1B activity, and not to thermal instability or proteolysis, because PTP-1B inactivated by prior treatment with vanadate had no effect (Fig. 1f). The temporal induction of Raf-1 activity by irradiation was noted with other cell types, including head and neck squamous-cell carcinoma, and breast carcinoma (U.K., unpublished results).

Tyrosine phosphorylation of Raf-1 could be attributed to the activation of protein-tyrosine kinase(s) or inhibition of a protein-tyrosine phosphatase(s). We examined the effect of radiation on protein-tyrosine phosphorylation and protein-tyrosine kinase activity in PCI-04A cells. The maximum stimulation of tyrosine phosphorylation of several protein bands was noted within 5 min of irradiation when we analysed cell lysates either by direct immunoblotting with anti-pTyr antibody or probing anti-pTyr immunoprecipitates with anti-pTyr antibody (Fig. 2a). We also found that protein-tyrosine kinase activity was increased ~3-fold within 5 min of irradiation (Fig. 2b). Additionally, irradiation stimulated apparent interaction of Raf-1 with unidentified ~140K and ~160K tyrosine-phosphorylated proteins which appeared later in anti-Raf-1 immunoprecipitates (Fig. 2c, top). As discussed earlier, no detectable change in the level of Raf-1

§ Present address: Department of Radiation Oncology, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

protein was noted after irradiation (Fig. 2c, bottom). Prior treatment with a protein-tyrosine kinase inhibitor, genistein, prevented the appearance of 140K and 160K proteins in anti-Raf-1 immunoprecipitates (Fig. 2c, top), and reduced (by 37%) the amount of Raf-1 protein immunoprecipitable with anti-pTyr antibody (Fig. 2d).

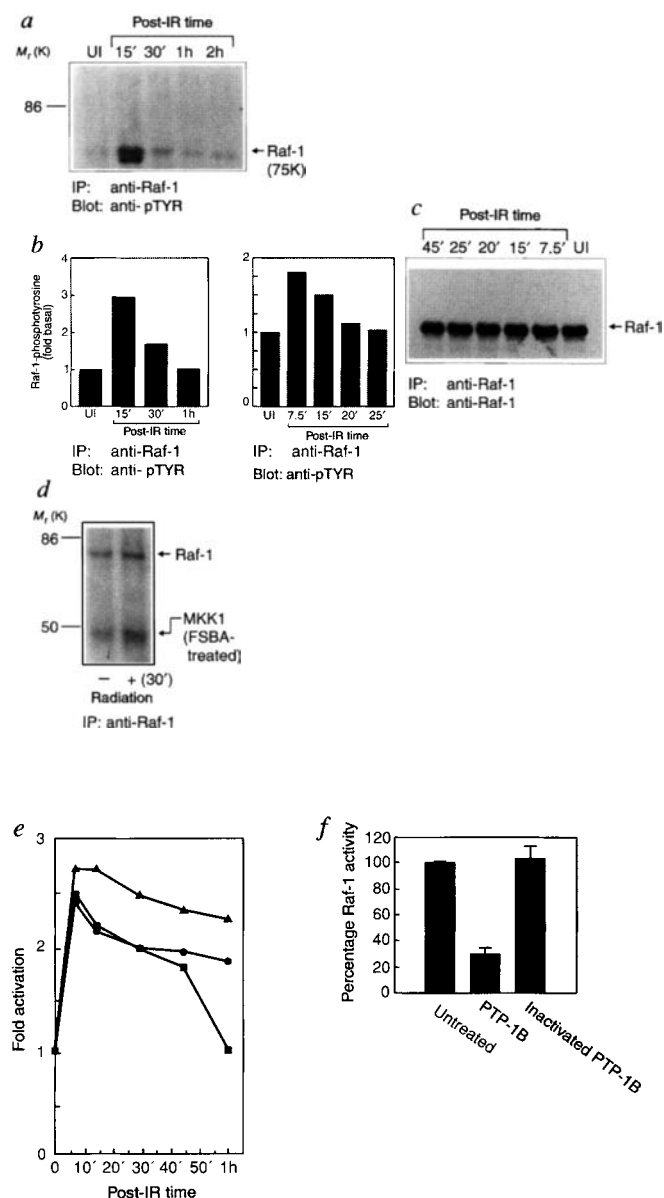
We next investigated whether radiation-stimulated tyrosine phosphorylation of Raf-1 occurred after transport of Raf-1 to the membrane. PCI-04A cells were fractionated into subcellular components at various times after irradiation. Fractionation was confirmed by immunoblotting of lysates with antibodies against a membrane marker (EGFR) or a predominantly cytoplasmic one (GST) (Fig. 3a). We then analysed membrane fractions for Raf-1 protein and Raf-1-phosphotyrosine and found a time-dependent increase in the amounts of each membrane-bound protein (Fig.

3b, c). For example, irradiation caused a fourfold increase in membrane-bound Raf-1 protein and a concurrent sixfold increase in Raf-1 phosphotyrosine within 7.5 min (Fig. 3d). Cytosolic fractions did not contain Raf-1-phosphotyrosine (data not shown). Probing Raf-1 immunoprecipitates with anti-pTyr antibody also revealed a gradual increase in two high-molecular-weight tyrosine-phosphorylated proteins (Fig. 3c).

The mechanisms by which Raf-1 is activated by ionizing radiation are incompletely understood. We have shown that Raf-1 is recruited to membranes and tyrosine-phosphorylated by radiation-responsive protein-tyrosine kinases. Several observations indicate that only the membrane-associated forms of Raf-1 are appreciably phosphorylated on tyrosine⁸⁻¹³. Specifically, Raf-1 immunoprecipitated from membranes of *v-ras*-transformed NIH3T3 cells is tyrosine-phosphorylated, whereas Raf-1 immu-

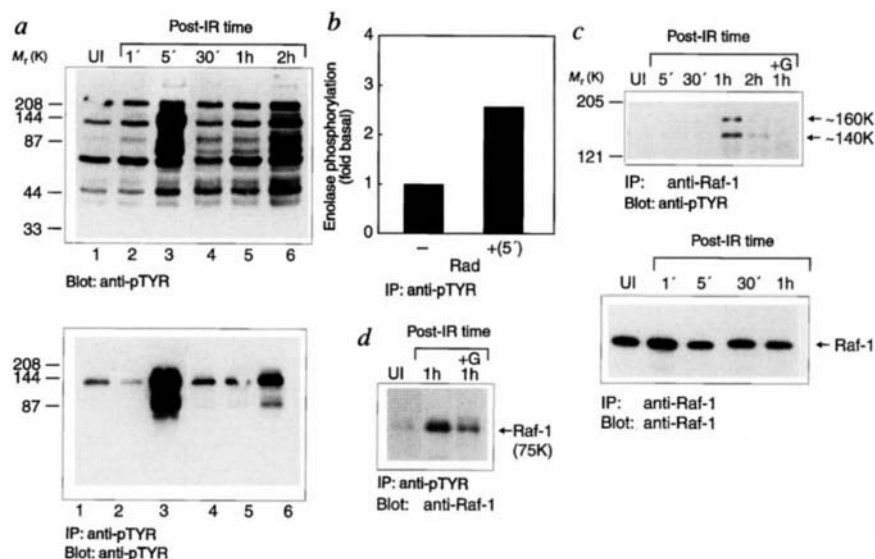
FIG. 1 Ionizing radiation (IR) stimulates tyrosine phosphorylation and enzymatic activity of Raf-1 protein kinase *in vivo*. **a**, Tyrosine phosphorylation of Raf-1. At the indicated times after irradiation of PCI-04A cells, Raf-1 was immunoprecipitated (IP) from cell lysates with agarose-conjugated anti-Raf-1 antibody (C-12, Santa Cruz) and probed by immunoblotting with monoclonal anti-pTyr antibody (Upstate Biotechnology) UI, unirradiated. **b**, Quantification of Raf-1 phosphotyrosine. Raf-1 immunoprecipitates were probed either with anti-pTyr antibody (experiment 1, left), or first with monoclonal anti-Raf-1 antibody (c-Raf-1, amino acids 162–378; Transduction Lab), followed by reprobing of the same blot with anti-pTyr antibody (experiment 2, right). **c**, Expression of Raf-1 protein in irradiated cells. Raf-1 was immunoprecipitated as in **a** and the immunoprecipitates probed with monoclonal anti-Raf-1 antibody. **d**, Stimulation of Raf-1 protein kinase activity. PCI-04A cells were lysed at the indicated time after irradiation. Raf-1 was immunoprecipitated from cell lysates with agarose-conjugated anti-Raf-1 antibody and its phosphotransferase activity assayed *in vitro* using a physiological substrate, MKK1 (refs 22–24). Radiolabelled reaction products were separated by electrophoresis and autoradiographed. FSBA, 5'-p-fluorosulphonylbenzoyl-adenosine. **e**, Stimulation of Raf-1, MKK and MAP kinase. PCI-04A cells were lysed at the indicated times after IR treatment. Enzymatic activities of Raf-1 (■), MKK (●) and MAP kinase (MAPK) (▲) from cells were measured individually by specific immune-complex kinase assays. Data were normalized as the fold increase (over control, unirradiated samples) of phosphate incorporation into substrate. 'Unstimulated' reference values (minus preimmune serum controls) are: Raf-1, 3.5 pmol into K52R (coupled assay); MKK, 8.5 pmol into K52R; MAPK, 20 pmol into myelin basic protein. Data representative of 3 experiments are shown. **f**, Raf-1 inactivation by PTP-1B. Samples were collected at 7.5 min after IR treatment. Data are mean $\pm$ s.d. from 3 experiments.

METHODS. PCI-04A cells were established from a human laryngeal squamous cell carcinoma²⁵ and were grown and maintained in culture (Z. Alatas *et al.*, unpublished). Logarithmically growing cells were subcultured (1:1) in complete medium containing 15% fetal bovine serum and incubated for 24 h. This was followed by serum starvation (17–21 h) before irradiation (15 Gy, 114 cGy min⁻¹). Irradiations were performed using a ¹³⁷Cs-gamma irradiator (J. L. Shepherd, Mark I series). For immunoprecipitation and immunoblotting, cells were lysed at the indicated times after irradiation in cold lysis buffer (500 mM HEPES, pH 7.2, 1% Nonidet P-40, 10% glycerol, 5 mM NaVO₄, 1 mM PMSF, 10 μ g ml⁻¹ aprotinin, 10 μ g ml⁻¹ leupeptin) and supernatants (whole-cell lysates) were collected after centrifugation (10,000g for 30 min). Cell lysates were normalized for protein concentration; 2–3 mg total cell protein was used for immunoprecipitation and immunoblotting with the indicated antibodies and enzyme-linked chemiluminescence (ECL, Amersham) as described^{7,12,13}. Quantifications by computer densitometry used Image-Quant software (Molecular Dynamics). In **d**, the MKK1 phosphorylation reaction was done in kinase buffer containing 30 mM HEPES, pH 7.4, 1 mM MnCl₂, 15 mM MgCl₂, 1 mM DTT, 0.1 mM ATP and 20 μ Ci [³²P]ATP (6,000 Ci mmol⁻¹) as described²⁶. In **e** and **f**, Raf-1, MKK and MAP kinase were individually immunoprecipitated from 0.5 mg lysate protein using immobilized anti-Raf-1 (Santa Cruz Biotechnology), anti-MKK, and anti-p42^{mapk} antibodies (gifts from M. Weber) and washed sequentially with buffer A (PBS, 0.1% (v/v) Triton X-100, 0.1% (v/v) 2-mercaptoethanol) containing 0.5 M LiCl, buffer A containing 0.05% (w/v) deoxycholate, and assay buffer (25 mM HEPES, pH 7.4, 15 mM MgCl₂, 0.1% (v/v) 2-mercaptoethanol). Raf-1 was assayed by MKK-coupled phosphorylation of K52R p42^{mapk} essentially as described¹². MKK and MAP kinase assays were single-stage assays using K52R and MBP²⁶. Raf-1 immune complexes on agarose beads were treated (at 37 °C for 1 h) with 1 μ g recombinant PTP-1B¹³ or the same amount of PTP-1B



inactivated by 0.5 mM vanadate. PTPase reactions were quenched by addition of 0.5 mM vanadate (37 °C, 10 min). Beads were recovered, supernatants discarded and Raf-1 activity measured by MKK-coupled phosphorylation of K52R p42^{mapk}¹². Additional control experiments omitting Raf-1 immune complexes and substituting constitutively active MKK verified that inactivation of PTP-1B by vanadate sufficed to prevent any interference with K52R phosphorylation in the assay (data not shown).

FIG. 2 Radiation stimulates protein-tyrosine phosphorylation and complex formation between Raf-1 and tyrosine-phosphorylated ~140K and ~160K proteins. **a**, Tyrosine phosphorylation of several unidentified proteins. Whole-cell lysates were prepared at the indicated times after irradiation (15 Gy) and used for direct immunoblotting with monoclonal anti-pTyr antibody (top panel). Cell lysates were also immunoprecipitated with agarose-conjugated monoclonal anti-pTyr antibody (Upstate Biotechnology) and the immunoprecipitates probed by blotting with monoclonal anti-pTyr antibody (bottom). **b**, Quantification of protein-tyrosine kinase activity. Tyrosine-phosphorylated proteins were immunoprecipitated with monoclonal anti-pTyr antibody and protein-tyrosine kinase activity in these immunoprecipitates was measured by phosphorylation of rabbit muscle enolase (Sigma). Reaction products were separated by electrophoresis, followed by autoradiography and quantification of the radiolabelled enolase band (results are representative of three experiments). **c**, **d**, Interaction of Raf-1 and tyrosine-phosphorylated ~140K and ~160K proteins *in vivo*. Cell lysates were prepared at the indicated times after irradiation (15 Gy). Normalized protein contents (2–3 mg) of various samples were used for immunoprecipitation with either monoclonal anti-Raf-1 antibody (**c**-Raf-1, amino acids 162–378; Transduction Labs; **c**, top), or monoclonal anti-pTyr antibody (**d**). Immunoprecipitates were then probed with monoclonal anti-pTyr antibody (**c**, top), or monoclonal anti-Raf-1 antibody (**d**). Expression of Raf-1 protein after



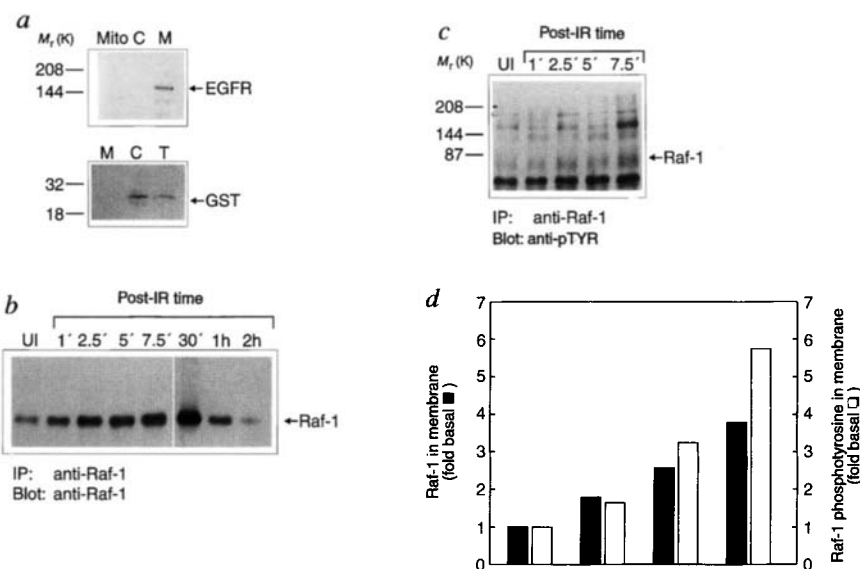
irradiation was monitored as described for Fig. 1c (**c**, bottom). +G₁, cells were treated with genistein (ICN Biomedicals; 100 μ g ml⁻¹, 1 h) before irradiation. Assays are described in Fig. 1 legend. Protein-tyrosine kinase activity was assayed as described²⁷.

noprecipitated from cytosolic fractions is not¹². This agrees with our data because Raf-1 phosphotyrosine was readily detected by western blotting of immunoprecipitates from 25 μ g of solubilized membranes from irradiated cells (data not shown), whereas detection with whole-cell lysates required milligram amounts of protein (Fig. 1a). Furthermore, Raf-1 immunoprecipitated from membranes of v-ras-transformed NIH3T3 cells can be inactivated by protein-tyrosine phosphatase 1B, suggesting that tyrosine phosphorylation is necessary for Raf-1 activation^{12,13}.

Alternative mechanisms of Raf-1 activation by irradiation treatment can be inferred from other work and are not excluded. Ionizing radiation (15 Gy) increases the level of Ras.GTP in breast

cancer cells (S.S. and U.K., unpublished data), consistent with the requirement for Raf activation. Ionizing radiation (10 Gy) increases the membrane content of ceramide in bovine endothelial cells¹⁴. Moreover, ceramide generated in response to tumour-necrosis factor(TNF)- α activates an unidentified 97K membrane-associated serine-threonine kinase, which in turn phosphorylates and activates Raf-1 (ref. 15). Irradiation (4–50 Gy) of endothelial cells stimulates synthesis of basic fibroblast growth factor (bFGF) and autocrine activation of protein kinase C (refs 16, 17); bFGF stimulates Raf-1 protein kinase¹⁸. Although ceramide production leads to apoptotic cell death¹⁴, it is interesting that bFGF protects against radiation-induced cell killing¹⁷.

FIG. 3 Membrane recruitment and tyrosine phosphorylation of Raf-1 in irradiated cells. **a**, Fractionation of PCI-04A cells into mitochondrial (Mito), membrane (M), and cytoplasmic (C) fractions. For control of the fractionation procedure, the receptor for epidermal growth factor (EGFR) served as a marker for membrane distribution and glutathione-S-transferase (GST) for cytoplasmic distribution. Lysates from unfractionated cells (total, T) and fractionated cells were normalized for protein content and probed by immunoblotting with monoclonal antibody against EGFR (Amersham; top). Similar aliquots were incubated with GST beads (Pharmacia), separated by 12.5% SDS-PAGE, and stained with Coomassie brilliant blue to visualize the expected 26K GST band (bottom). **b**, Membrane recruitment of Raf-1. Membrane fractions of PCI-04A cells were isolated at the indicated times after irradiation (15 Gy). Lysates were normalized for protein contents, and Raf-1 protein was then detected by immunoprecipitation with agarose-conjugated anti-Raf-1 antibody, followed by blotting of these immunoprecipitates with monoclonal anti-Raf-1 antibody. Results were similar in two independent experiments. **c**, Tyrosine phosphorylation of Raf-1 in the membrane. Membrane-associated Raf-1 phosphotyrosine was identified by probing of Raf-1 immunoprecipitates with monoclonal Raf-1 antibody, and then reprobing of the same blot with



monoclonal anti-pTyr antibody. **d**, Quantification of membrane-bound total and tyrosine-phosphorylated Raf-1. Assays are described in Fig. 1 legend. Cell fractionation has been described⁹.

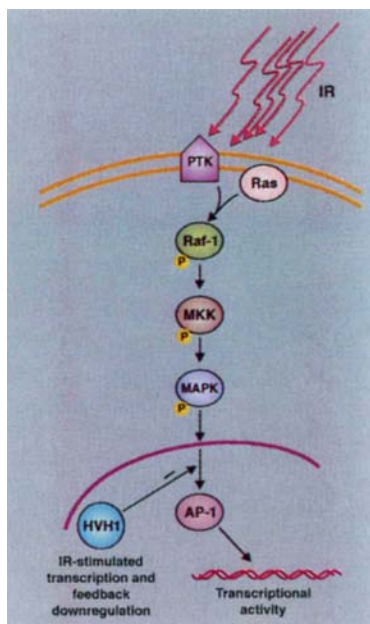


FIG. 4 Model of IR-responsive signal transduction pathway. Radiation activates protein-tyrosine kinase(s) (PTK) and Ras, leading to the phosphorylation and activation of Raf-1. Activated Raf-1 phosphorylates its known physiological substrate MKK, which in turn results in downstream activation of the transcription factor AP-1 through MAPK. Radiation also induces expression of HVH1, an MAPK-specific dual phosphatase^{28,29}. An increased level of HVH1 may dephosphorylate MAPK, resulting in the negative feedback regulation of MAPK and AP-1 activities.

Thus, ionizing radiation, which lacks the specificity of ligand-receptor interactions, may activate multiple Raf-1-activation pathways, which then activate MAP kinase¹⁹ (Fig. 1e). It is not known which of these pathways predominates, but increasing evidence points to the importance of tyrosine phosphorylation in the Raf-1 activation mechanism⁷⁻¹³. In PCI-04A cells, activation of the MAP kinase pathway by irradiation leads to activation of transcription factor AP-1 and induction of a MAP-kinase-specific dual phosphatase HVH1 (U.K. *et al.*, unpublished data). Based on our results, a model for irradiation-induced signalling response is proposed (Fig. 4).

Our data are consistent with the idea that Raf-1 is an important substrate of irradiation-stimulated tyrosine kinase. The induction of an irradiation signalling pathway (Fig. 4) may play a key role in the repair of damage by a newly synthesized cellular constituent(s) of the repair system. This is supported by our earlier observation of a correlation between antisense cDNA-mediated downregulation of Raf-1, delayed tumour growth and enhanced radiosensitivity³, and is consistent with the ultraviolet-light-mediated signal transduction cascade involving tyrosine kinases, Ha-Ras, Raf-1 and Jun^{20,21}. □

Received 1 April; accepted 27 June 1996.

- Kolch, W., Heidecker, G., Lloyd, P. & Rapp, U. *Nature* **349**, 426–428 (1991).
- Williams, N. G. & Roberts, T. M. *Cancer Metast. Rev.* **13**, 105–116 (1994).
- Kasid, U. *et al. Science* **243**, 1354–1356 (1989).
- Kasid, U., Pirollo, K., Dritschilo, A. & Chang, E. *Adv. Cancer Res.* **61**, 195–233 (1993).
- Uckun, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 252–256 (1993).
- Hallahan, D. E. *et al. J. Biol. Chem.* **268**, 4903–4907 (1993).
- Fabian, J. R., Daar, I. O. & Morrison, D. K. *Molec. cell. Biol.* **13**, 7170–7179 (1993).
- Leevers, S. J., Paterson, H. F. & Marshall, C. J. *Nature* **369**, 411–414 (1994).
- Traverse, S. *et al. Oncogene* **8**, 3175–3181 (1993).
- Stokoe, D., MacDonald, S. G., Cadwallader, K., Symons, M. & Hancock, J. F. *Science* **264**, 1463–1467 (1994).
- Marais, R., Light, Y., Paterson, H. F. & Marshall, C. J. *EMBO J.* **14**, 3136–3145 (1995).
- Jelinek, T., Dent, P., Sturgill, T. W. & Weber, M. J. *Molec. cell. Biol.* **16**, 1027–1034 (1996).
- Dent, P., Jelinek, T., Morrison, D. K., Weber, M. J. & Sturgill, T. W. *Science* **268**, 1902–1906 (1995).
- Haimovitz-Friedman, A. *et al. J. exp. Med.* **180**, 525–535 (1994).
- Yao, B. *et al. Nature* **378**, 307–310 (1995).

- Witte, L. *et al. Cancer Res.* **49**, 5066–5072 (1989).
- Haimovitz-Friedman, A. *et al. Cancer Res.* **51**, 2552–2558 (1991).
- Morrison, D. K., Kaplan, D. R., Rapp, U. R. & Roberts, T. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8855–8859 (1988).
- Kharbanda, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 5416–5420 (1994).
- Devary, Y., Gottlieb, R. A., Smeal, T. & Karin, M. *Cell* **71**, 1081–1091 (1992).
- Radler-Pohl, A. *et al. EMBO J.* **12**, 1005–1012 (1993).
- Dent, P. *et al. Science* **257**, 1404–1407 (1992).
- Howe, L. R. *et al. Cell* **71**, 335–342 (1992).
- Kyriakis, J. M. *et al. Nature* **358**, 417–421 (1992).
- Heo, D. S. *et al. Cancer Res.* **49**, 5167–5175 (1989).
- Dent, P. *et al. Biochem J.* **303**, 105–112 (1994).
- Chae, H. P., Jarvis, L. J. & Uckun, F. M. *Cancer Res.* **53**, 447–451 (1993).
- Zheng, C.-F. & Guan, K.-L. *J. Biol. Chem.* **268**, 16116–16119 (1993).
- Kwak, S. P., Hakes, D. J., Martell, K. J. & Dixon, J. E. *J. Biol. Chem.* **269**, 3596–3604 (1994).

ACKNOWLEDGEMENTS. We thank F. Uckun and D. Morrison for discussion; A. Dritschilo for his support; D. Massaro for access to the Image-Quant computer program for densitometry; F.-H. Wang for laboratory management; and L. Culp for artwork. This work was supported by a Shannon Award and a grant from the NIH (U.K.), and the Howard Hughes Medical Institute (T.W.S.). P.D. is a fellow of the Juvenile Diabetes Foundation International.

CORRESPONDENCE and requests for materials should be addressed to U.K. (e-mail: kasidu@gunet.georgetown.edu).

Cooperation of the tumour suppressors IRF-1 and p53 in response to DNA damage

Nobuyuki Tanaka*†, Masahiko Ishihara*, Marc S. Lamphier*‡, Hiroaki Nozawa*, Toshifumi Matsuyama§, Tak W. Mak§, Shinichi Aizawa||, Takashi Tokino¶, Moshe Oren# & Tadatsugu Taniguchi*

* Department of Immunology, Faculty of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan

† Nara Institute of Science and Technology, Takayama 8916-5, Ikomashi, Nara 630-1, Japan

‡ Precursory Research Program for Embryonic Science and Technology, Research Development Corporation of Japan, Hikari-dai 1-7, Seika-cho, Kyoto 619-02, Japan

§ Amgen Institute, Ontario Cancer Institute, and Department of Immunology and Medical Biophysics, University of Toronto, Ontario M4X 1K9, Canada

|| Institute of Molecular Embryology and Genetics, Kumamoto University School of Medicine, Honjo 2-2-1, Kumamoto 860, Japan

¶ Laboratory of Molecular Medicine, Institute of Medical Science, University of Tokyo, Shirogane-dai 4-6-1, Minato-ku, Tokyo 108, Japan

Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

NORMALLY growing cells promptly cease DNA synthesis when exposed to genotoxic stresses, such as radiation, and this cell-cycle arrest prevents the accumulation of mutations^{1,2}. The transcription factor interferon regulatory factor (IRF)-1 is essential for the regulation of the interferon system³⁻⁵, inhibits cell growth, and manifests tumour-suppressor activities^{6,7}. Here we show that mouse embryonic fibroblasts (EFs) lacking IRF-1 are deficient in their ability to undergo DNA-damage-induced cell-cycle arrest. A similar phenotype has been observed in EFs lacking the tumour suppressor p53 (refs 8, 9), although the expression of IRF-1 and p53 are independent of one another. Furthermore, we show that transcriptional induction of the gene encoding p21 (*WAF1*, *CIP1*)¹⁰⁻¹², a cell-cycle inhibitor, by γ -irradiation is dependent on both p53 and IRF-1, and that the p21 promoter is activated, either directly or indirectly, by both in a transient cotransfection assay. These two tumour-suppressor transcription factors therefore converge functionally to regulate the cell cycle through the activation of a common target genes.

We have previously observed that EFs lacking IRF-1 are susceptible to oncogenic transformation by activated c-Ha-ras, and differ from wild-type EFs in their apoptotic response to DNA damage⁶. To investigate further the role of IRF-1 in the regulation

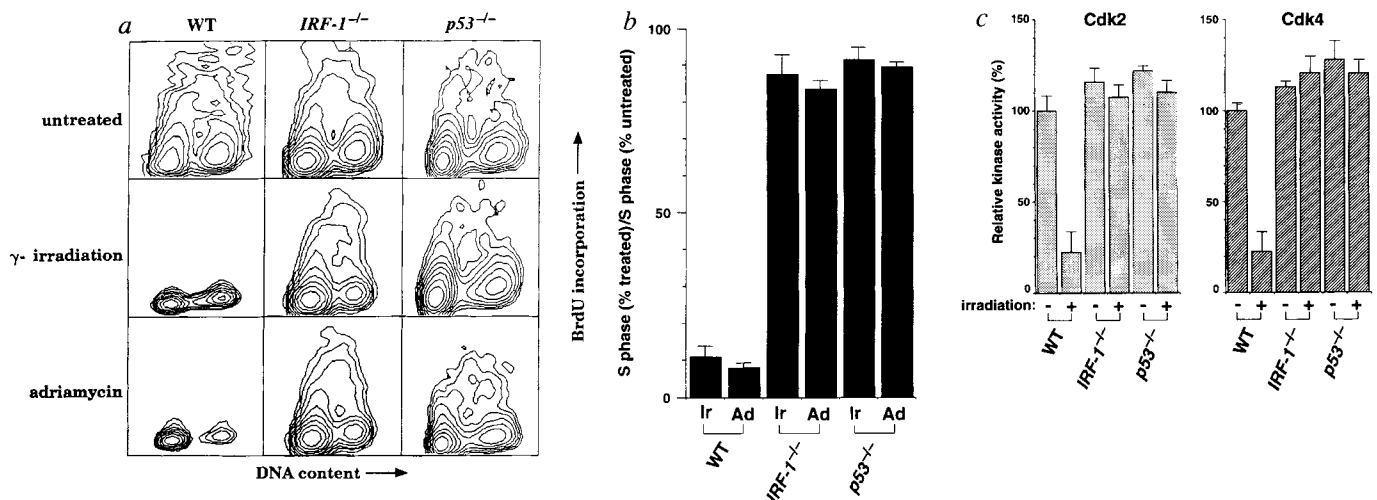


FIG. 1 DNA damage-induced cell-cycle arrest and CDK activities of *IRF-1*^{-/-} and *p53*^{-/-} EFs. **a**, Representative two-dimensional flow-cytometric analysis after γ -irradiation or adriamycin treatment. Asynchronous cultures of wild-type (WT), *IRF-1*^{-/-} and *p53*^{-/-} EFs were subjected to γ -irradiation at a dose of 20 Gy, or treated with 0.5 μ g ml⁻¹ adriamycin. The DNA synthesis rate was evaluated 24 h after treatment with a 30-min BrdU pulse and staining with a fluorescein isothiocyanate-conjugated anti-BrdU antibody. Cells in the G1 and G2/M phase of the cell cycle appear in the bottom left and right quadrants, respectively. Cells in S phase are positive for BrdU, and are detected in the arch that connects the G1 and G2/M populations. **b**, Quantitative analysis of the number of cells entering S phase 24 h after γ -irradiation (Ir) or treatment with adriamycin (Ad), compared with the number of cells in untreated cells. **c**, Activities of Cdk2 (left) and Cdk4

(right) kinase after γ -irradiation. Cdk2 or Cdk4 immunoprecipitates from either untreated or irradiated EFs (as indicated) were prepared 10 h after γ -irradiation (20 Gy). Precipitates were tested for kinase activities using histone H1 for Cdk2 or retinoblastoma protein (Rb) for Cdk4. In **b** and **c**, histograms show the mean of two independent experiments; error bars, s.d. **METHODS**. EFs were prepared from wild-type, *IRF-1*^{-/-} and *p53*^{-/-} mice as described⁶, and grown in DMEM with 10% fetal calf serum. Cells were plated at a density of 1×10^6 cells per 10-cm dish, and irradiated 24 h after plating. Cell-cycle analysis was as described²⁰, with a FACScan using CellQuest software (Becton-Dickinson). Immunoprecipitation and kinase assays were done as described²⁷. Kinase activities were quantified using a Fujix BAS 2000 image analyser.

of the cell cycle, EFs were isolated from mice carrying a null mutation in both *IRF-1* alleles (*IRF-1*^{-/-} EFs), treated with γ -irradiation, and subjected to cell-cycle analysis; wild-type EFs and EFs lacking p53 (*p53*^{-/-} EFs) were treated similarly. Wild-type EFs were substantially arrested in phases G1 and G2/M of the cell cycle 24 h after γ -irradiation (Fig. 1a), and the number of S-phase cells was ~90% lower in irradiated than in non-irradiated cells (Fig. 1b). In contrast, the number of *IRF-1*^{-/-} EFs in S phase was only ~10% lower after γ -irradiation, a value similar to that obtained with *p53*^{-/-} EFs¹³ (Fig. 1b). Similar results were also observed following treatment with the anti-cancer drug adriamycin (Fig. 1a, b). These results indicate that both IRF-1 and p53 are essential for DNA damage-induced cell-cycle arrest.

It has been demonstrated that cell-cycle arrest in response to DNA damage is mediated by the inhibition of G1 cyclin-dependent kinases (CDKs), such as Cdk2 and Cdk4 (refs 2, 14). An *in vitro* kinase assay for Cdk2 and Cdk4 indicated that these activities were strongly inhibited in wild-type EFs after γ -irradiation, but that inhibition was virtually abrogated in *IRF-1*^{-/-} EFs, again showing similarity to *p53*^{-/-} EFs¹⁵ (Fig. 1c). However, IRF-1 differs from p53 in that it is also involved in the regulation of the interferon (IFN) system, playing an essential role in the host defence against viral and bacterial infections^{4,5}. The IFN response requires the transcriptional upregulation of *IRF-1*, which is dependent on the factor p91 (STAT1)¹⁶, and IFN-induced cell-cycle arrest also seems to require p91 (ref. 17). However, although deletion of p91 eliminated *IRF-1* induction by IFN^{18,19}, it had no effect on radiation-induced cell-cycle arrest (data not shown). Thus, the role of IRF-1 in radiation-induced cell-cycle arrest seems to be distinct from its role in the IFN response.

The results of the kinase assay suggested that the function of one or more of the inhibitors of G1 cyclin/CDK complexes, the CKIs^{2,14}, may be defective in *IRF-1*^{-/-} EFs. We therefore examined the expression of several CKI genes by RNA blotting analysis.

It was found that basal expression of *p21* mRNA was dramatically reduced in *IRF-1*^{-/-} and *p53*^{-/-} EFs (Fig. 2a). Long exposure using an image analyser revealed weak expression in both *IRF-1*^{-/-} and *p53*^{-/-} EFs (at least 10-fold lower than in wild-type EFs) and no induction after γ -irradiation (data not shown). The role of IRF-1 and p53 in the radiation-induced transcriptional activation of the *p21* gene was also studied in a nuclear run-on assay, which demonstrated that *p21* transcription was induced upon γ -irradiation of wild-type EFs but not *IRF-1*^{-/-} or *p53*^{-/-} EFs (data not shown). Consistent with these observations, p21 expression was increased in irradiated wild-type EFs, but not in *IRF-1*^{-/-} or *p53*^{-/-} EFs (Fig. 2b). However, the expression of mRNAs for other known CKI genes, such as *p27*, *p15*, *p16*, *p18* and *p19* (refs 2, 14), was not affected in either *IRF-1*^{-/-} or *p53*^{-/-} EFs either before or after γ -irradiation (data not shown).

We also examined whether the expression of IRF-1 and p53 occur independently of one another. As shown in Fig. 2c, γ -irradiation caused a two- to threefold increase in IRF-1 protein levels, whereas no induction of *IRF-1* transcription was observed (data not shown). The same treatment also caused a four- to fivefold increase in p53 protein levels²⁰, and these inductions of IRF-1 and p53 occurred in the absence of one another (Fig. 2c). Thus IRF-1 and p53 seem to cooperate through two parallel and independent pathways in the induction of cell-cycle arrest and *p21* gene transcription.

It has been well documented that the induction of *p21*, a dual inhibitor of CDKs and proliferating-cell nuclear antigen (PCNA), is indeed essential for DNA damage-induced cell-cycle arrest, and that *p21* induction by DNA damage is dependent on p53 (ref. 21). The *p21* promoter sequence has been analysed extensively, and evidence has been provided showing that p53 binds to and activates the promoter²². In addition, one MyoD binding site and three IFN-activated GAF (for γ -activated factor¹⁶) binding sites have been found within the promoter, and these sites seem to be activated in a p53-independent manner^{17,23,24}. When the con-

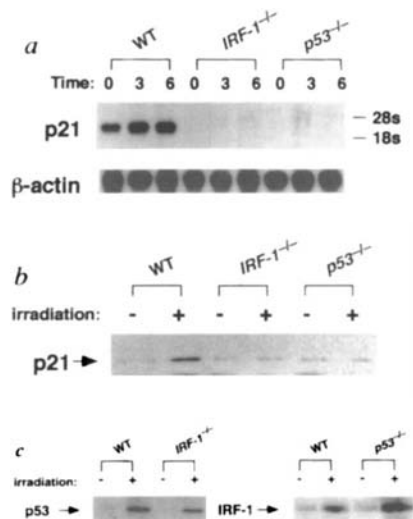


FIG. 2 Expression of p21 in *IRF-1*^{-/-} and *p53*^{-/-} EFs. **a**, Expression of p21 mRNAs were analysed by RNA blotting analysis of RNA isolated from wild-type, *IRF-1*^{-/-} and *p53*^{-/-} EFs untreated and after irradiation (20 Gy), and quantified using a Fujix BAS 2000 image analyser; times (in hours) after irradiation are indicated. The same filter was reprobbed with β -actin cDNA. **b**, Western blot of p21 protein. Whole-cell lysates from untreated and irradiated EFs were analysed by immunoblotting using anti-p21 antibody. **c**, Induction of p53 protein in *IRF-1*^{-/-} EFs, and IRF-1 protein in *p53*^{-/-} EFs, after γ -irradiation. Whole-cell lysates were isolated from wild-type, *IRF-1*^{-/-} and *p53*^{-/-} EFs untreated and 1 h after irradiation (20 Gy), and analysed by western blotting using anti-p53 or anti-IRF-1 monoclonal antibodies. **METHODS**. Total RNA (5 μ g) was prepared by guanidinium thiocyanate/CsCl centrifugation and analysed by RNA blotting as described⁶. Immunodetection was performed as described⁶, with a rabbit anti-mouse p21 antibody (Santa Cruz, sc#-471), monoclonal anti-p53 antibody (Oncogene Research, Pab 240) and monoclonal anti-IRF-1 antibody²⁸.

tribution of IRF-1 and p53 in the activation of the *p21* promoter was examined by a transient cotransfection assay, the *p21* promoter was found to be activated by IRF-1 or p53, and activation was further augmented by coexpression of the two factors (Fig. 3a). The p53-mediated activation of an artificial promoter containing multiple p53 sites but no IRF-1 site (IRF-E) was not affected by cotransfection with IRF-1, so it is unlikely that IRF-1 acts through p53 (Fig. 3b). Examination of the *p21* promoter showed that there are three potential IRF-1 binding sites (see Fig. 3a legend). However, in this transient cotransfection assay we were unable to observe any consistent effect of deleting these sites on the cooperative activation of the *p21* promoter by p53 and IRF-1 (N.T., M.I. and H.N., unpublished observation). Thus it is not yet clear whether IRF-1 acts directly or indirectly to activate the *p21* promoter; further work is required to clarify this point.

IRF-1 and p53 possess both non-overlapping and overlapping functions in the cell. For example, as mentioned above, IRF-1 is required for antiviral and antibacterial responses^{4,5}, but p53 is not. Each factor is also essential for DNA damage-induced apoptosis in distinct lymphocyte populations⁷. Furthermore, p53 controls the mitotic spindle checkpoint²⁵, but this function is normal in *IRF-1*^{-/-} EFs (N.T., unpublished observation). However, both p53 and IRF-1 are essential for the DNA-damage-induced cell-cycle arrest of normal EFs (see above) and apoptosis of EFs carrying an activated Ha-ras oncogene⁶. Mutations in the *p53* gene have been detected in a broad range of human cancers, although so far the involvement of *IRF-1* in human cancer has not been extensively studied, except in the case of some haematopoietic neoplasms²⁶. Thus it will be interesting to analyse the role of IRF-1 in other human cancers, and to investigate whether it regulates other target genes in common with p53. □

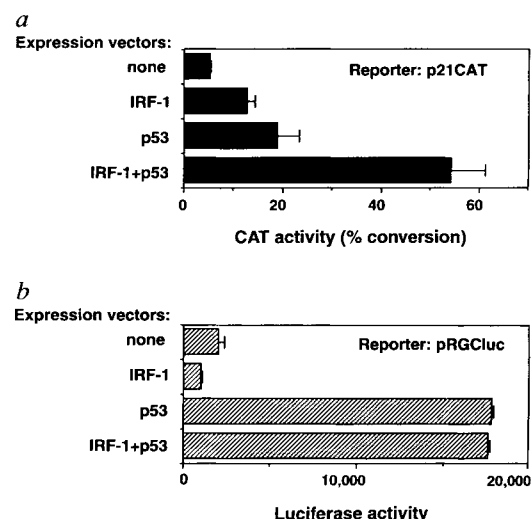


FIG. 3 Transient cotransfection analysis of IRF-1 and p53. **a**, *p21* promoter activities from cells cotransfected with the *p21*CAT reporter gene (containing the promoter region of the mouse *p21* gene up to base pair -4661 relative to the TATA box)²², along with IRF-1 (pAct1) (ref. 4) and p53 (pC53-SN3) (ref. 29) expression vectors. Three potential IRF-Es were found in the mouse *p21* promoter at -2255 (CAAAGTAAACA), -2237 (TAGAAGT-GACATT) and -1817 (CAAAGTAAACA). *p21*CAT (5 μ g) was cotransfected with 2.5 μ g pAct1 and/or 0.01 μ g pC53-SN3, or the same amount of control plasmid, pActC, by the calcium phosphate-DNA precipitation method⁴, and CAT enzyme activity was analysed as described²². The assay was repeated four times; results were essentially reproducible. This activation by IRF-1 and p53 was abrogated when a reporter construct deleted for the -4661 to -1781 upstream region (containing two p53 binding sites²²) of the *p21* promoter was assayed. **b**, pRGLuc reporter gene³⁰ (5 μ g) was cotransfected with 1 μ g pC53-SN3 and/or 4 μ g pAct1, or an amount of pActC to bring the total to 5 μ g, and luciferase enzyme activity was analysed by Lumat LB9501 (Berthold). The assay was repeated twice; results were essentially reproducible. In **a** and **b**, histograms show the mean of two independent experiments; error bars, s.d.

Received 20 March; accepted 23 June 1996.

- Hartwell, L. *Cell* **71**, 543–546 (1992).
- Hunter, T. & Pines, J. *Cell* **79**, 573–582 (1994).
- Matsuyama, T. *et al.* *Cell* **75**, 83–97 (1993).
- Kamijo, R. *et al.* *Science* **263**, 1612–1615 (1994).
- Kimura, T. *et al.* *Science* **264**, 1921–1924 (1994).
- Tanaka, N. *et al.* *Cell* **77**, 829–839 (1994).
- Tamura, T. *et al.* *Nature* **376**, 596–599 (1995).
- Vogelstein, B. & Kinzler, K. W. *Cell* **70**, 523–526 (1992).
- Liebermann, D. A., Hoffman, B. & Steinman, R. A. *Oncogene* **11**, 199–210 (1995).
- El-Deiry, W. S. *et al.* *Cell* **75**, 817–825 (1993).
- Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. *Cell* **75**, 805–816 (1993).
- Xiong, Y. *et al.* *Nature* **366**, 701–704 (1993).
- Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. *Cell* **74**, 957–967 (1993).
- Sherr, C. J. & Roberts, J. M. *Genes Dev.* **9**, 1149–1163 (1995).
- Brugarolas, J. *et al.* *Nature* **377**, 552–557 (1995).
- Darnell Jr, J. E., Kerr, I. M. & Stark, G. R. *Science* **264**, 1415–1421 (1994).
- Chin, Y. E. *et al.* *Science* **272**, 719–722 (1996).
- Meraz, M. A. *et al.* *Cell* **84**, 431–442 (1996).
- Durbini, J. E., Hackenmiller, R., Simon, M. C. & Levy, D. E. *Cell* **84**, 443–450 (1996).
- Kastan, M. B., Onyekwure, O., Sidransky, D., Vogelstein, B. & Craig, R. W. *Cancer Res.* **51**, 6304–6311 (1991).
- El-Deiry, W. S. *et al.* *Cancer Res.* **54**, 1169–1174 (1994).
- El-Deiry, W. S. *et al.* *Cancer Res.* **55**, 2910–2919 (1995).
- Parker, S. B. *et al.* *Science* **267**, 1024–1027 (1995).
- Halevy, O. *et al.* *Science* **267**, 1018–1021 (1995).
- Cross, S. M. *et al.* *Science* **267**, 1353–1356 (1995).
- Willman, C. L. *et al.* *Science* **259**, 968–971 (1993).
- Matsushima, H. *et al.* *Molec. Cell Biol.* **14**, 2066–2076 (1994).
- Harada, H. *et al.* *Science* **259**, 971–974 (1993).
- Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. V. & Vogelstein, B. *Science* **249**, 912–915 (1990).
- Haupt, Y., Rowan, S., Shaulian, E., Voudsen, K. H. & Oren, M. *Genes Dev.* **9**, 2170–2183 (1995).

ACKNOWLEDGEMENTS. We thank R. D. Schreiber, D. Beach, C. J. Sherr and B. Vogelstein for p91 / mice, mouse *p21* and *p16* cDNA, mouse *p18* and *p19* cDNA, and pC53-SN3, respectively; Z.-J. Liu for technical advice; and Y. Taya for advice and discussion. This work was supported in part by a special grant for Advanced Research on Cancer from the Ministry of Education, Science and Culture of Japan, and by Yamanouchi foundation SKK.

CORRESPONDENCE and requests for materials should be addressed to T. Tanaka.

Two distinct actions of retinoid-receptor ligands

Jia-Yang Chen, John Clifford, Chris Zusi*, John Starrett*, David Tortolani*, Jacek Ostrowski*, Peter R. Reczek*, Pierre Chambon & Hinrich Gronemeyer

Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, Collège de France, BP 163, 67404 Illkirch Cedex, CU de Strasbourg, France

* Bristol-Myers Squibb Pharmaceutical Research Institute, 100 Forest Avenue, Buffalo, New York 14213-1091, USA

SIGNALLING by all-*trans* retinoic acid is mediated through RXR-RAR retinoid receptor heterodimers^{1,2}, in which RXR has been considered to act as a transcriptionally silent partner³⁻⁵. However, we show here that in cultured NB4 (ref. 6) human acute promyelocytic leukaemia⁷⁻⁹ cells treated with either an RAR- α -selective agonist alone, or certain RAR- α antagonists in combination with an RXR agonist, receptor-DNA binding is induced *in vivo*, resulting in expression of the target genes of retinoic acid as well as acute promyelocytic leukaemia protein (PML) relocation to nuclear bodies¹⁰⁻¹² and differentiation before apoptosis. These results indicate that RAR- α ligands can induce two separate events: one enables RXR-RAR- α heterodimers to bind to DNA *in vivo* and allows RXR agonists to act; the other induces transcriptional activity of RAR- α . The availability of receptor-specific synthetic retinoids that can induce distinct receptor functions has potential in extending the therapeutic repertoire of retinoids.

NB4 (ref. 6) cells carry the t(15;17) chromosomal translocation that is a feature in most cases of acute promyelocytic leukaemia (APL) and generates a PML-RAR- α fusion protein which causes a differentiation block⁷⁻⁹. We exposed these cells to receptor-selective synthetic retinoids¹³ (Table 1a) in order to investigate the

contributions of RAR- α /PML-RAR- α , RAR- β , RAR- γ and RXRs to the molecular and cellular events leading to cellular differentiation upon treatment with all-*trans* retinoic acid (*t*-RA). Am80 (RAR- α -specific at ≤ 1 nM; ref. 13), BMS753 (a pure RAR- α agonist) and BMS681 (an RAR- α , β agonist and RAR- γ antagonist) efficiently induced differentiation (Table 1b), whereas retinoids lacking RAR- α agonist activity (BMS453 and BMS411 (ref. 13), BMS961, BMS614, and the RXR agonist SR11237 (ref. 14), also known as BMS649 (ref. 15, 16)), were ineffective on their own (Table 1b and Fig. 1a-e). Moreover, the differentiation induced by 1 nM Am80 was completely blocked by an excess of the pure RAR- α antagonist BMS614 (not shown). These results show that inducing the transcriptional activity of the non-fused RAR- α allele and/or the RAR- α moiety of PML-RAR- α is sufficient to relieve the differentiation block¹⁷ resulting from the formation of a PML-RAR- α fusion protein. Note that the phenomenon of AP-1 transrepression by RAR- α is unlikely to be critical for retinoic acid-induced NB4 cell differentiation, because 'dissociated' retinoids such as BMS453, although they efficiently repress AP-1 activity through all three RARs¹³, did not induce differentiation on their own (Fig. 1c).

After 4 days of treatment with *t*-RA or RAR- α agonists, fluorescence-activated cell sorting (FACS) analysis showed that NB4 cells had reduced mitotic activity and accumulated in the G1/G0 phases of the cell cycle (Fig. 2b, e; compare with untreated cells (Fig. 2a)). On day 6, sub-2N particles accumulated in *t*-RA-treated cultures (Fig. 2c) and on day 8, extensive DNA fragmentation occurred that was indicative of apoptosis (Fig. 2d). As in the case of differentiation, only RAR- α (but not RAR- β or RAR- γ) agonists could induce this sequence of events (Table 1b, Fig. 2f and data not shown).

Before differentiation, as reported previously¹⁰⁻¹² for *t*-RA, RAR- α agonists caused association of endogenous PML/PML-RAR- α with nuclear bodies (Fig. 1g), whereas RAR- α antagonists did not (Fig. 1h, and data not shown). DNA footprinting *in vivo* showed that RAR- α agonists, but not antagonists, induced the occupation of the prototype retinoic-acid-response element (RARE) and of an adjacent DR5-RARE¹⁸ present in the

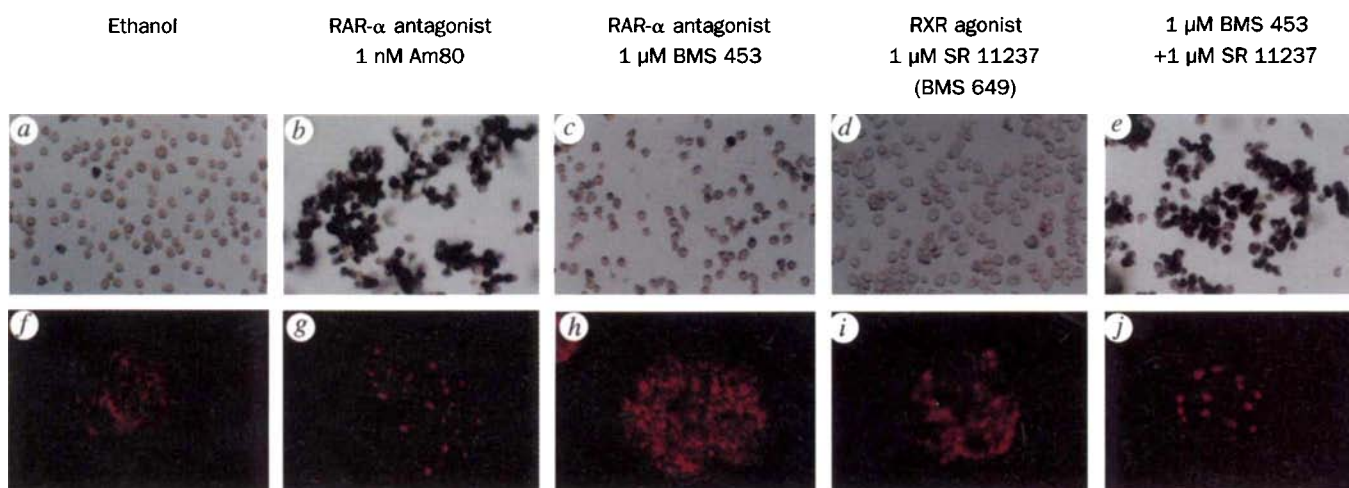


FIG. 1 Synergy between RAR- α agonists or antagonists and RXR agonists for the induction of NB4-cell differentiation and PML relocation to nuclear bodies. **a-e**, Nitroblue tetrazolium (NBT) staining of NB4 cells treated for 4 days with ethanol or the indicated retinoid(s). **f-j**, Confocal microscopy images of anti-PML immunofluorescence staining of NB4 cells treated for 4 d with the indicated retinoid(s).

METHODS. **a-e** For NBT staining, NB4 cells were grown in RPMI1640 (plus 2 mM glutamine), containing 100 U μ l⁻¹ penicillin and streptomycin, and 8% FCS. Cells were treated for 4 d with ethanol or retinoid(s), washed and resuspended at a density of 5×10^5 cells per ml. 50- μ l aliquots of this suspension were spread over slides coated with poly-L-lysine (Sigma). After

washing with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 1.4 mM KH₂PO₄ pH 7.3), cells were fixed for 10 min in 2% formalin (37% formaldehyde in 15% methanol)/PBS at room temperature. After 3 washes in PBS, cells were scored by TPA (12-O-tetradecanoyl-phorbol-13-acetate; 200 ng ml⁻¹)-induced NBT reduction (Sigma fast BCIP/NBT) for 30 min at 37 °C. Samples were examined with an optical microscope (Diavert Leitz) and the percentage of differentiated cells determined by counting at least 300 cells for each treatment. **f-j**, Anti-PML immunofluorescence staining of NB4 cells was done essentially as described¹¹ using polyclonal anti-PML antibodies (from A. Dejean) and donkey anti-rabbit IgG (H + L) conjugated with cyanine 3 (Cy3) (Jackson Labs).

TABLE 1 NB4 cell differentiation by synthetic retinoids

(a) Retinoid characteristics										
Receptor	BMS retinoid									
	753	453	411	961	681	614				
RAR- α	+	—	—	0	+	—				
RAR- β	0	+	+	(+)	+	0				
RAR- γ	0	—	(+)	+	—	0				

(b) Percentage of cells differentiated										
Retinoid (nM)							+100 nM SR11237 (BMS649)			
	t-RA	9c-RA	Am80	BMS 753	BMS 681	Am80	BMS 753	BMS 681	BMS 453	BMS 614
0.1	—	—	<1	—	—	<1	<1	<1	—	—
0.3	—	—	<1	—	—	20	<1	<1	—	—
0.5	—	—	<1	—	<1	40	<1	5	—	—
1	<1	<1	80	<1	<1	80	5	5	<1	<1
10	<1	4	81	1	3	—	45	80	65	3
100	81	65	80	85	75	—	80	80	78	5
1,000	80	70	—	90	80	—	90	85	—	10

						SR11237 (BMS649)				
	BMS 453	BMS 411	BMS 961	BMS 614						
1,000	<1	<1	<1	<1		<2				

a, RAR- α , β or γ agonistic (plus signs) or antagonistic (minus signs) activities were determined by transient transfection and reporter cell analysis as described¹³. (+), Weak agonistic activity; 0, absence of agonistic and antagonistic activities. b, Percentage of NB4 cells differentiated as determined by NBT staining (Fig. 1).

RAR β 2 promoter¹⁸, as well as recruiting other promoter-binding factors (Fig. 3a). These footprints were almost indistinguishable from those for mouse P19 embryonal carcinoma cells¹⁹ (Fig. 3b). Am80-induced expression of RAR- β 2 (and RAR- γ) was confirmed by polymerase chain reaction with reverse transcription (RT-PCR); there was no induction with RAR- α antagonists alone (BMS453 in Fig. 3c, and results not shown).

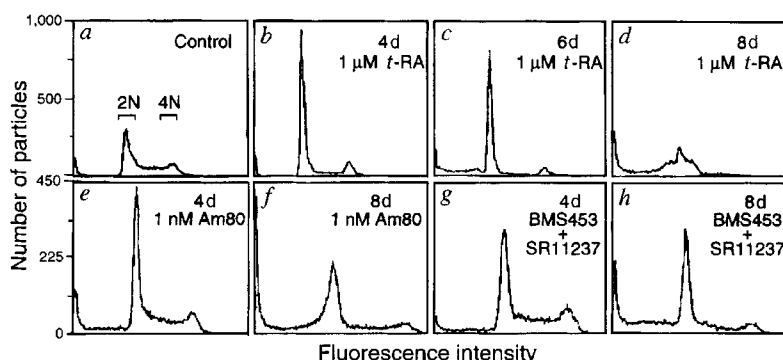
We investigated the function of RXR in these reactions beyond its role as a heterodimerization partner, by testing whether RAR- and RXR-specific ligands could synergize. In the presence of RXR-specific SR11237, both Am80 and BMS753 were significantly more effective at inducing differentiation at low concentrations (Table 1b; SR11237 alone did not induce differentiation: Fig. 1d). Strikingly, in combination with SR11237, even the RAR- α antagonist BMS453 induced differentiation more effectively than *t*-RA or 9-*cis* retinoic acid (9c-RA) (Table 1b, Fig. 1e). It is unlikely that this differentiation could be due to autoinduction of RAR- β 2 by RAR- β agonist activity¹³ (Table 1a) of BMS453 in

the presence of SR11237, because BMS411 (Table 1a), a stronger RAR- β agonist than BMS453 (ref. 13), did not synergize with SR11237 to induce RAR- β 2 promoter occupancy (Fig. 3a, b), RAR- β 2 messenger RNA synthesis, or differentiation (results not shown). In addition, BMS614, a 'pure' RAR- α antagonist which does not bind RAR- β or RAR- γ , also induced differentiation in the presence of SR11237, in contrast to the RAR- γ (β) agonist BMS961, which does not bind RAR- α (Table 1, and data not shown). The lower efficiency of BMS614 relative to BMS453 may be related to its lower binding affinity for RAR- α .

Together, these results indicate that RAR- α (or PML-RAR- α) is specifically involved in the synergistic induction of differentiation in the presence of the RXR agonist. Synergy between RAR- α and RXR ligands was not only observed for differentiation, but also for subsequent apoptosis, as shown by the strong antiproliferative and apoptotic effect of a combination of the RAR- α antagonist BMS453 with the RXR agonist SR11237, both of which were ineffective by themselves (Fig. 2g, h). BMS453/SR11237-

FIG. 2 Effect of retinoids on cell-cycle distribution and appearance of sub-2N apoptotic cells and particles. a, Number of untreated NB4 cells containing a 2N, 4N or an intermediate complement of DNA. About 20,000 particles are represented in each panel. The effect of 1 μ M *t*-RA treatment for 4, 6 and 8 d (b–d), 1 nM Am80 for 4 and 8 d (e, f; note that in f the position of the peak corresponds to 2N particles), and a combination of 100 nM BMS453 and 100 nM SR11237 for 4 and 8 days (g, h). Forward- and side-scatter analysis of the same cell preparations, as well as the presence of apoptotic DNA 'ladders', confirmed these results (data not shown).

METHODS. Cell-cycle distribution and presence of sub-2N cells and particles in control and retinoid-treated NB4 cells were determined by cell-cycle flow cytometry based on DNA content, using an Epics Profile II cell sorter (Coulter Electronics) equipped with a 15W argon laser at a wavelength of 588 nm and an emission wavelength of 575 nm. Cultures of untreated or retinoid-treated cells were centrifuged (250g), and fixed in 70% ethanol and stored at -20°C . After two washes in PBS, cells and subcellular particles were incubated in 1 mg ml⁻¹ RNase A (59 Kunitz U per mg; Sigma) for 30 min at 37 $^{\circ}\text{C}$, centrifuged and resuspended



in PBS at $\sim 10^6$ cells per ml. Ethidium bromide was added to 50 $\mu\text{g ml}^{-1}$ immediately before sample analysis.

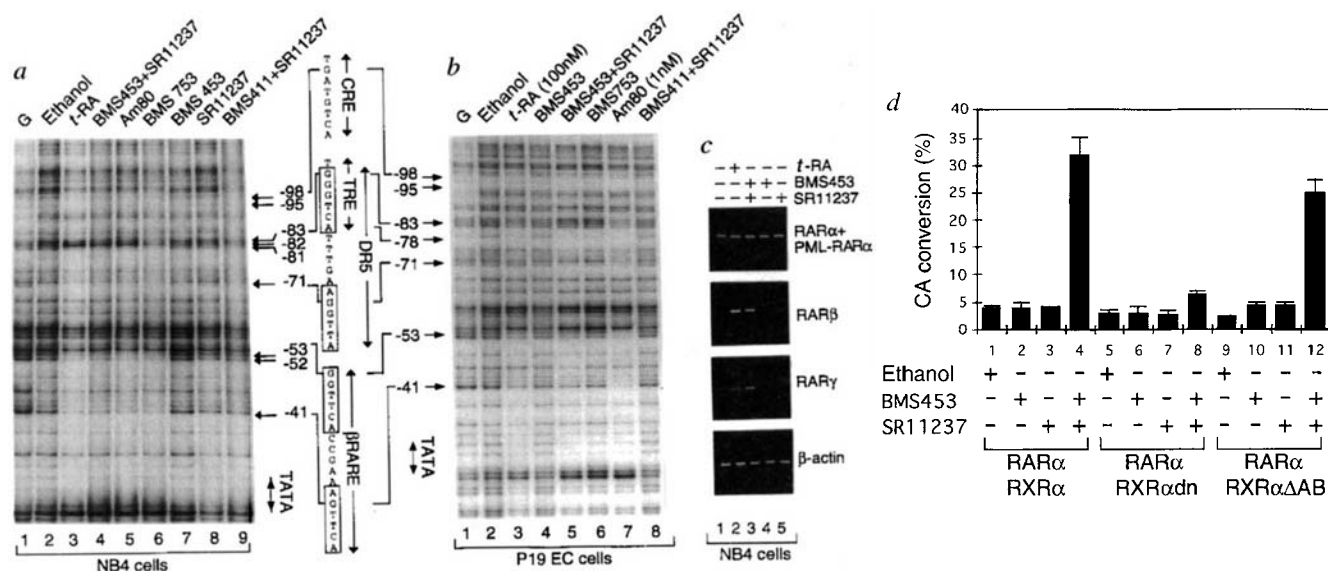


FIG. 3 RAR- α and RXR synergize to induce a dimethyl sulphate (DMS) genomic footprint over the RAR- β 2 promoter region in NB4 or P19 cells and RAR- β and RAR- γ mRNA production. NB4 (a) or P19 (b) cells were treated for 24 h with ethanol or 100 nM *t*-RA, 100 nM BMS453 plus 100 nM SR11237, 100 nM BMS753, 1 nM Am80, 1 μ M BMS453, 1 μ M SR11237, or 1 μ M BMS411 plus 1 μ M SR11237, as indicated. Cells were treated with DMS, and DNA was purified and digested by piperidine. After precipitation with ethanol, samples were processed for ligation-mediated PCR (LM-PCR) and resolved on a 4.8% sequencing gel. *In vitro* controls (G ladders) were produced by reacting purified DNA with DMS. c, PCR with reverse transcription (RT-PCR) determination of RAR- α , β , and γ transcripts in NB4 cells treated for 24 h with either ethanol vehicle, 100 nM *t*-RA, 100 nM BMS453 plus 100 nM SR11237, 1 μ M BMS453, or 1 μ M SR11237 (lanes 1 to 5, respectively). RNA transcripts of the β -actin gene were used as a control. Samples were resolved on 6% non-denaturing

polyacrylamide gels. d, RXR- α AF-2 is required for BMS453/SR11237 synergy in transiently transfected P19 EC cells. Note that synergy is seen with RAR- α wild-type or N-terminally deleted (RXR α ΔAB) RXR- α but not with a mutant (RXR α dn)²² in which the C-terminal AF-2 AD (refs 1, 2) has been removed.

METHODS. P19.6 and NB4 cells were treated with either ethanol or the indicated retinoid(s) for 24 h. After washing in PBS, cells were exposed to 0.1% dimethylsulphate (Aldrich) for 5 min at room temperature. High-molecular-weight DNA was extracted and cleaved with piperidine. DMS treatment *in vitro* of naked DNA was done as described²⁶. LM-PCR was done according to ref. 27, except that Deep Vent (exo⁻) DNA polymerase (BioLabs) was used. RT-PCR was carried out according to standard procedures. Transient transfections were done as described¹³ using 200 ng RAR- α and RXR- α wild-type or mutant expression vectors^{22,28}, 5 μ g RAR- β 2-CAT reporter²² and 1 μ M synthetic retinoids.

induced differentiation and apoptosis was preceded by an association of PML-RAR- α with nuclear bodies which was indistinguishable from that seen in the presence of RAR- α agonists (Fig. 1g,j, compare with untreated cells, f). Moreover, occupation of the RAR- β 2 RARE and transcription factor recruitment¹⁹ to the RAR- β 2 promoter *in vivo* was also induced when cells were treated with BMS453 and SR11237 (Fig. 3a) and led to an efficient accumulation of RAR- β (and RAR- γ) mRNA (Fig. 3c). In keeping with its inability to synergize with SR11237 for differentiation, the other RAR- α antagonist/RAR- β agonist BMS411 was not able to induce nuclear body association of PML-RAR- α (not shown) or RAR- β 2 promoter occupancy (Fig. 3a).

This synergism between RAR and RXR ligands is difficult to reconcile with previous reports^{3,4} that RXR is a transcriptionally silent partner in RXR-RAR heterodimers on both DR1 and DR5 RAREs (reviewed in ref. 5). Specifically, how could a heterodimer comprising an antagonist-bound RAR- α and an agonist-bound RXR bind to and activate transcription from the RAR- β 2 promoter, if RXR were unable to bind its ligand when associated with DNA? One report⁴ distinguishes between the inhibition by ligand-free and ligand-bound RAR on RXR activity. Moreover, in contrast to the earlier work^{3,4} *in vitro*^{20,21} and ligand-binding data with purified RAR-RXR DNA complexes (J. Iyer and H.G., unpublished results) indicate that both partners of RXR-RAR heterodimers can bind their cognate ligand irrespective of their association with DNA, supporting our conclusion that both partners can be transcriptionally active²². Regardless of the conflicting *in vitro* results, it is clear that *in vivo* the RXR partner of RXR-RAR dimers can respond to an agonist ligand and therefore must bind to it, which is in keeping with the *in vivo* synergism between RAR and RXR agonists^{15,16,23,24}.

Our results show that RAR agonists at high concentration are by themselves sufficient to induce the genetic programmes leading to differentiation and apoptosis of NB4 cells, as well as transcription from the RAR- β 2 promoter. In contrast, an RXR agonist is inactive unless it is associated with an RAR agonist or certain RAR antagonists. It appears therefore that (1) the apo-RAR structure in the RXR-RAR heterodimer is incompatible with events that are necessary for transactivation *in vivo*, and (2) transactivation by an RXR-RAR heterodimer comprises two separate RAR-mediated events: one for binding of RXR-RAR- α heterodimers to DNA *in vivo* which allows ligand-dependent RXR activation of AF-2 (refs 1, 2) by agonists and can be induced by either an RAR- α agonist or certain RAR- α antagonists; the other induces RAR- α AF-2 activity and can be exerted by RAR- α agonists only. The AF-2 function of RXR must mediate synergism between RAR and RXR ligands because in transient transfection experiments, BMS453/SR11237 synergy was observed with AF-1-deleted (RXR α ΔAB) and wild-type, but not with AF-2-disabled (RXR α dn) RXRs (Fig. 3d). These results agree with the observation that RXR mutants devoid of AF-2 activity decrease *t*-RA-induced promoter occupancy and RAR- β 2 expression in stable transfectants²³.

Neither RAR- α specificity nor the synergism between RAR- α /BMS453 and RXR/SR11237 were specific for NB4 cells or PML-RAR- α . Human HL60 myeloblastic leukaemia cells lacking the PML-RAR- α fusion protein responded like NB4 cells to the various ligand combinations with respect to RAR- β 2 promoter occupancy, RAR- β 2 mRNA accumulation and differentiation (not shown). But, in contrast to NB4 cells, apoptosis of HL60 cells cannot be achieved with an RAR agonist alone, but needs an RXR agonist²⁵. RAR- β 2 promoter occupancy and RAR- β 2

mRNA accumulation is also seen in P19 embryonal carcinoma (EC) cells treated with a combination of an RAR- α antagonist and an RXR agonist¹⁵ (Fig. 3b). In contrast, the same ligand combination in F9 EC cells did not induce either RAR- β 2 expression or differentiation¹⁶. Considered together with the ability of 'dissociated' retinoids¹³ to repress AP1 activity, our results indicate that it may be possible to initiate complex gene programmes in a cell-specific manner by the appropriate choice/combination of synthetic retinoids with predefined characteristics. The selectivity as well as the synergistic potential of RAR and RXR ligands, which allow much-reduced concentrations of the individual compounds to achieve biological activity, hold promise for therapeutic application. □

Received 7 May; accepted 5 July 1996.

1. Gronemeyer, H. & Laudet, V. *Protein Profile* **2**, 1173–1308 (1985).
2. Chambon, P. *FASEB J.* (in the press).
3. Kurokawa, R. et al. *Nature* **371**, 528–531 (1994).
4. Forman, B. M., Umesono, K., Chen, J. & Evans, R. M. *Cell* **81**, 541–550 (1995).
5. Mangelsdorf, D. J. & Evans, R. M. *Cell* **83**, 835–850 (1995).
6. Lanotte, M. et al. *Blood* **77**, 1080–1086 (1991).
7. Warrel, R. P. et al. *N. Engl. J. Med.* **329**, 177–189 (1993).
8. Grignani, F. et al. *Blood* **83**, 10–25 (1994).
9. Lavau, C. & Dejean, A. *Leukemia* **8**, 1615–1621 (1994).

10. Dyck, J. A. et al. *Cell* **76**, 333–343 (1994).
11. Weis, K. et al. *Cell* **76**, 345–358 (1994).
12. Koken, M. H. et al. *EMBO J.* **13**, 1073–1083 (1994).
13. Chen, J.-Y. et al. *EMBO J.* **14**, 1187–1197 (1995).
14. Lehmann, J. M. et al. *Science* **258**, 1944–1946 (1992).
15. Roy, B., Taneja, R. & Chambon, P. *Mol. Cell. Biol.* **15**, 6481–6487 (1995).
16. Taneja, R. et al. *Proc. Natl Acad. Sci. USA* **93**, 6197–6202 (1996).
17. Grignani, F. et al. *Cell* **74**, 423–431 (1993).
18. Valcarcel, R., Holz, H., Jimenez, C. G., Barettoni, D. & Stunnenberg, H. G. *Genes Dev.* **8**, 3068–3079 (1994).
19. Dey, A., Minucci, S. & Ozato, K. *Mol. Cell. Biol.* **14**, 8191–8201 (1994).
20. Kersten, S., Dawson, M. I., Lewis, B. A. & Noy, N. *Biochemistry* **35**, 3816–3824 (1996).
21. Apfel, R. et al. *J. Biol. Chem.* **270**, 30765–30772 (1995).
22. Durand, B. et al. *Cell* **71**, 73–85 (1992).
23. Blanco, J. C. G. et al. *Genes to Cells* **1**, 209–211 (1996).
24. Clifford, J. et al. *EMBO J.* (in the press).
25. Nagy, L. et al. *Mol. Cell. Biol.* **15**, 3540–3551 (1995).
26. Mueller, P. R. & Wold, B. *Science* **246**, 780–786 (1989).
27. Garrity, P. A. & Wold, B. J. *Proc. Natl Acad. Sci. USA* **89**, 1021–1025 (1992).
28. Nagpal, S., Friant, S., Nakshatri, H. & Chambon, P. *EMBO J.* **12**, 2349–2360 (1993).

ACKNOWLEDGEMENTS. We thank A. Dejean for anti-PML antibody, M. Lanotte for NB4 cells, T. Grange for protocols for DNA preparation and LM-PCR, Y. Lutz for immunocytochemistry, J.-L. Vonesch for recording confocal microscopy images, and the services for DNA synthesis, cell culture, photography and artwork. This work was supported by funds from the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Centre Hospitalier Universitaire Régional, the Ministère de la Recherche et de la Technologie, and the Collège de France. J.C. holds a fellowship from the US NIH.

CORRESPONDENCE and requests for materials should be addressed to H.G. (e-mail: hg@titus.u-strasbg.fr).

Dimeric ligands define a role for transcriptional activation domains in reinitiation

Steffan N. Ho*, Stephen R. Biggar*,
David M. Spencer*, Stuart L. Schreiber†
& Gerald R. Crabtree*

* Departments of Pathology and Developmental Biology, Howard Hughes Medical Institute, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305, USA
† Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, USA

EUKARYOTIC transcriptional activators mediate transcriptional induction through stabilization of the preinitiation complex^{1–4}, probably through direct interactions with basal transcription factors^{5–7}. *In vitro* studies on the role of an activator in the maintenance of on-going transcription (reinitiation) have been contradictory, suggesting that, after formation of a preinitiation complex, an activator may⁸ or may not⁹ be necessary for transcription to be maintained. We have developed a means of regulating transcription in living cells through the use of both homodimeric and heterodimerizing synthetic ligands that allow the ligand-dependent association and disassociation of a transcriptional activation domain with a promoter. Here we report that maintaining the transcription of endogenous genes *in vivo*, in both yeast and human cells, requires the continuous presence of the activation domain. The use of synthetic ligands as a transcriptional on-off switch represents a powerful means of controlling the transcription *in vitro* and *in vivo* for both experimental and therapeutic purposes.

Transcription factors are composed of functionally distinct domains that independently mediate DNA binding and transcriptional activation^{10,11}. The lack of a requirement for covalent association of these distinct domains¹² suggests that transcription factors could be made inducible through the use of a dimeric ligand that mediates the physical approximation of the separated domains and thereby activates transcription (Fig. 1a).

To associate functional domains of proteins, we made use of the lipid-soluble dimeric ligand FK1012, a dimeric form of FK506,

which directs interactions between proteins linked to the ligand receptor FKBP12 (ref. 13). To determine whether FK1012 could regulate transcription through the inducible recruitment of a transcriptional activation domain to a promoter, chimaeric proteins containing FKBP12 were constructed (Fig. 1a). Addition of FK1012 to cells transfected with both the DNA-binding domain (either GF3 or HF3) and the activation domain (NF3V1) constructs resulted in dose-dependent transcriptional activation with a half-maximal effective concentration (EC₅₀) of approximately 10 nM. Addition of the ligand had no effect on cells transfected with the reporter gene alone, or with the reporter gene plus either GF3, HF3 or NF3V1 alone (Fig. 1b, c). Transcriptional induction mediated by the GAL4 zinc-finger DNA-binding domain was similar to that mediated by the HNF1 homeodomain DNA-binding domain, indicating that the FK1012-dependent interaction between the DNA-binding domain and the transcriptional activation domain is not dependent on a specific DNA-protein conformation. The level of induction mediated by the GAL4 and the HNF1 DNA-binding domain construct was $35.4 \pm 4.5\%$ ($n = 6$) and $13.4 \pm 2.4\%$ ($n = 3$), respectively, of that observed in cells transfected with an optimized amount of either the GAL4-VP16 or HNF1-VP16 chimaera, in which the DNA-binding domain and the activation domain are directly linked.

The use of dimeric ligands, such as FK1012, to induce transcription suggests that monomeric forms of the ligand should disrupt FK1012-driven interactions and eliminate the association between the DNA-binding and activation domains (Fig. 1a). Incubation of transfected cells with both FK1012 and a 10-fold excess of the monomeric ligand (FK506-M) resulted in at least 90% inhibition of transcription (Fig. 2a). These results suggest that the addition of excess monomeric ligand blocks the functional association of the transcriptional activation domain with the DNA-binding domain, as DNA binding itself is not affected by ligand, as assessed by gel-shift analysis (data not shown).

To determine whether a transcriptional activation domain is required for the maintenance of on-going transcription *in vivo*, monomeric ligand was added at varying times after initiating transcription with dimeric ligand. Excess monomeric ligand resulted in a marked reduction in the rate of accumulation of secreted alkaline phosphatase (SEAP) (Fig. 2b) with a latency period following addition of monomeric ligand that is consistent with the rapid (within minutes) termination of transcription. The ligand (dimeric or monomeric) had no effect on either constitutive transcription mediated by GAL4-VP16, or inducible transcrip-

tion of a stably integrated NFAT (for nuclear factor of activated T-cells) responsive reporter gene (data not shown).

To measure directly the rate of transcription in the presence and absence of an activation domain, nuclear run-on experiments were performed using transfected Jurkat cells. Addition of FK1012 for 9 h resulted in a 6–10-fold increase in the rate of transcription of the SEAP gene. Subsequent addition of excess monomeric ligand resulted in a $64 \pm 3\%$ ($n = 3$) reduction in the induced rate of transcription of the SEAP gene within 30 min (Fig. 2c), a reduction similar to the half-life of 15–25 min for the dissociation of FK506 from FKBP12 (S.L.S., unpublished data). These results indicate that displacement of an activation domain from the promoter results in the rapid termination of transcription. However, these experiments were performed using transiently transfected plasmids, and therefore minimize the effects of chromatin structure and DNA methylation that might be involved in the transcriptional control of an integrated gene. We therefore examined cells with a stably integrated reporter gene, and also found that FK1012-induced transcription was rapidly blocked by addition of the monomeric ligand (Fig. 2d).

To study further the role of an activation domain in the maintenance of transcription *in vivo*, we investigated an integrated *GAL1-lacZ* gene, as well as the endogenous, single-copy *GAL1* gene of *Saccharomyces cerevisiae*. These experiments used ligands which, in monomeric form, mediate the heteromeric interaction between two distinct ligand-binding domains. One such ligand,

FK506, forms a complex with FKBP12 that interacts with calcineurin^{14,15}, and can thus function as a heterodimerizing ligand by mediating the ligand-dependent interaction between two distinct ligand-binding domains, FKBP12 and calcineurin A. Furthermore, the binding of FK506 to FKBP12 can be blocked by rapamycin, a ligand structurally similar to FK506 which competes for binding to FKBP12 but does not interact with calcineurin¹⁶. Rapamycin can therefore be used to interfere with or dissociate the FK506-mediated recruitment of an activation domain to a promoter. In a manner similar to FK506, rapamycin can also function as a heterodimerizing ligand by mediating the ligand-dependent interaction between FKBP12 and the FKBP-rapamycin binding (FRB) domain of FRP/RAFT1 (ref. 17).

The ability of FK506 to function as a heterodimerizing ligand to mediate the recruitment of an activation domain to an integrated promoter was demonstrated by measuring transcription from an integrated *GAL1-lacZ* gene. Addition of FK506 resulted in the dose-dependent transcriptional induction of β -galactosidase (Fig. 3a). This induction was dependent on the expression of both a GAL4 DNA-binding domain–FKBP12 (GAL4BD–hFKBP12) chimera and a calcineurin A–activation domain chimera. Similarly, rapamycin induced transcription in cells expressing GAL4BD–hFKBP12 and the FRB–activation domain chimera GAL4AD–FRB (Fig. 3b). As expected, induction of transcription by FK506 was blocked by simultaneous addition of excess rapamycin (Fig. 3c).

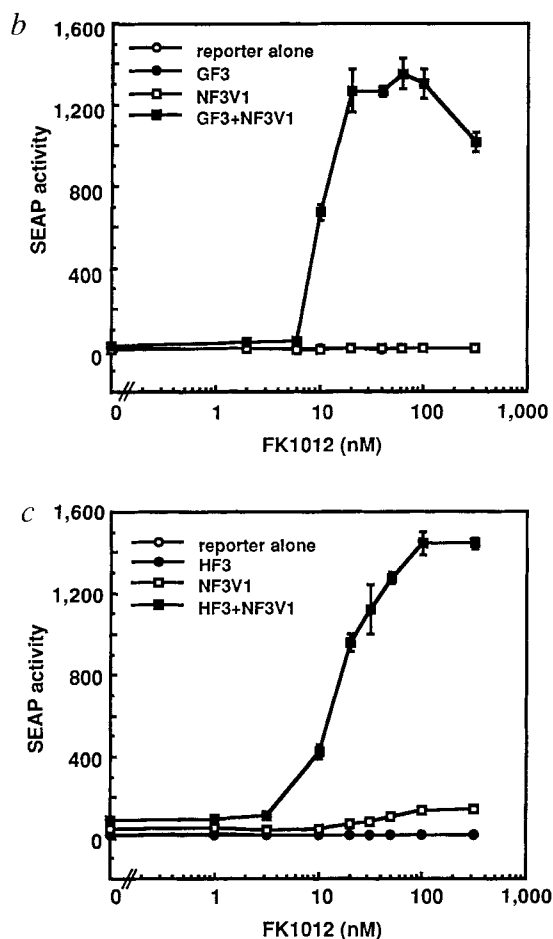
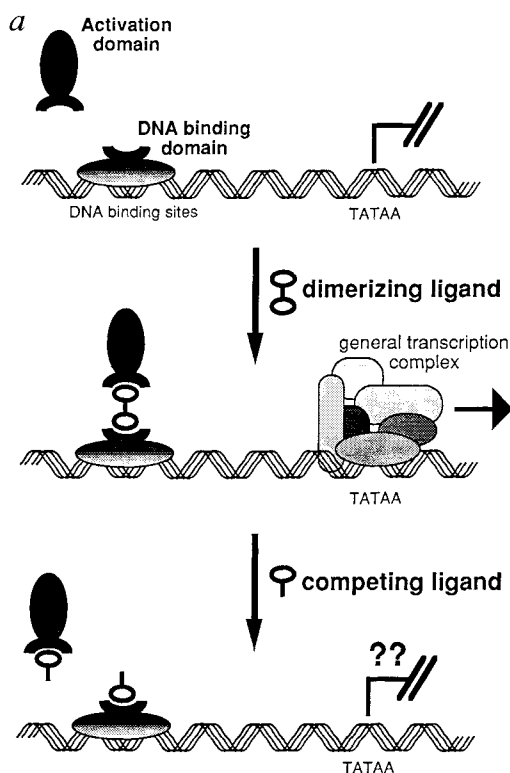


FIG. 1 Controlling transcription with synthetic ligands. *a*, Schematic representation of the recruitment of an activation domain to a DNA-binding domain with a dimeric ligand, and dissociation of the activation domain with the monomeric ligand. *b*, The reporter plasmid β 28IL-2SX (4 μ g) was transfected alone or with GF3 (1 μ g) and/or NF3V1 (10 μ g). The symbols

representing transfection of the reporter plasmid alone, reporter plus GF3, and reporter plus NF3V1, overlaid one another. *c*, The reporter plasmid β 28IL-2SX (4 μ g) was transfected alone or with HF3 (1 μ g) and/or NF3V1 (10 μ g). The symbols representing transfection of the reporter plasmid alone and reporter plus HF3 overlaid one another.

Displacing the activation domain with rapamycin after the induction of transcription by FK506 resulted in a marked reduction in β -galactosidase expression from the integrated gene (Fig. 4a). The effect of activation-domain dissociation was measured directly by analysing transcription of the endogenous *GAL1* gene. Given the short half-life of the *GAL1* RNA (~ 5 min)¹⁸, the amount present accurately reflects the rate of transcription of the *GAL1* gene. FK506-mediated recruitment of the GAL4 activation domain to either the endogenous *GAL1* promoter or to an integrated *GAL1-HIS3* gene resulted in the induction of transcription within 30 min (Fig. 4b, left). Subsequent addition of rapamycin to cells that had been induced to transcribe *GAL1* with FK506 resulted in a marked reduction in transcription within 20 min (Fig. 4b, right).

These results indicate that the continued presence of a transcriptional activation domain is required for the maintenance of on-going transcription *in vivo* in both mammalian and yeast cells. Steps subsequent to the formation of the preinitiation complex, in which transcription is maintained through reinitiation, differ from initiation in that an uncharacterized postinitiation complex remains at

the promoter, increasing the efficiency of transcription¹⁹. *In vitro* studies demonstrate differences between the rates of initiation and reinitiation²⁰, as well as differences in the requirement for RNA polymerase elongation factor II (ref. 21), indicating that the mechanism of transcription reinitiation differs from that of initiation. *In vitro* studies have shown that, after the initiation of transcription, TFIID remains bound to the promoter while other general transcription factors are released, including TFIIB^{22,23}. This reassociates with TFIID, reforming a site for the association of RNA pol II (refs 22, 23). *In vivo* studies in yeast suggest that the binding of TBP can be the rate-limiting step for transcription, and that an activator is not necessary for steps subsequent to the binding of TBP^{24,25}. Thus the requirement for an activator to maintain on-going transcription demonstrated here *in vivo* may reflect a requirement for the continuous stable association of TFIID with the promoter, which allows the repeated recruitment of the general transcription factors that dissociate from the complex following initiation of transcription.

Note added in proof: Rapamycin has also recently been utilized as a heterodimerizing agent in mammalian cells³¹. □

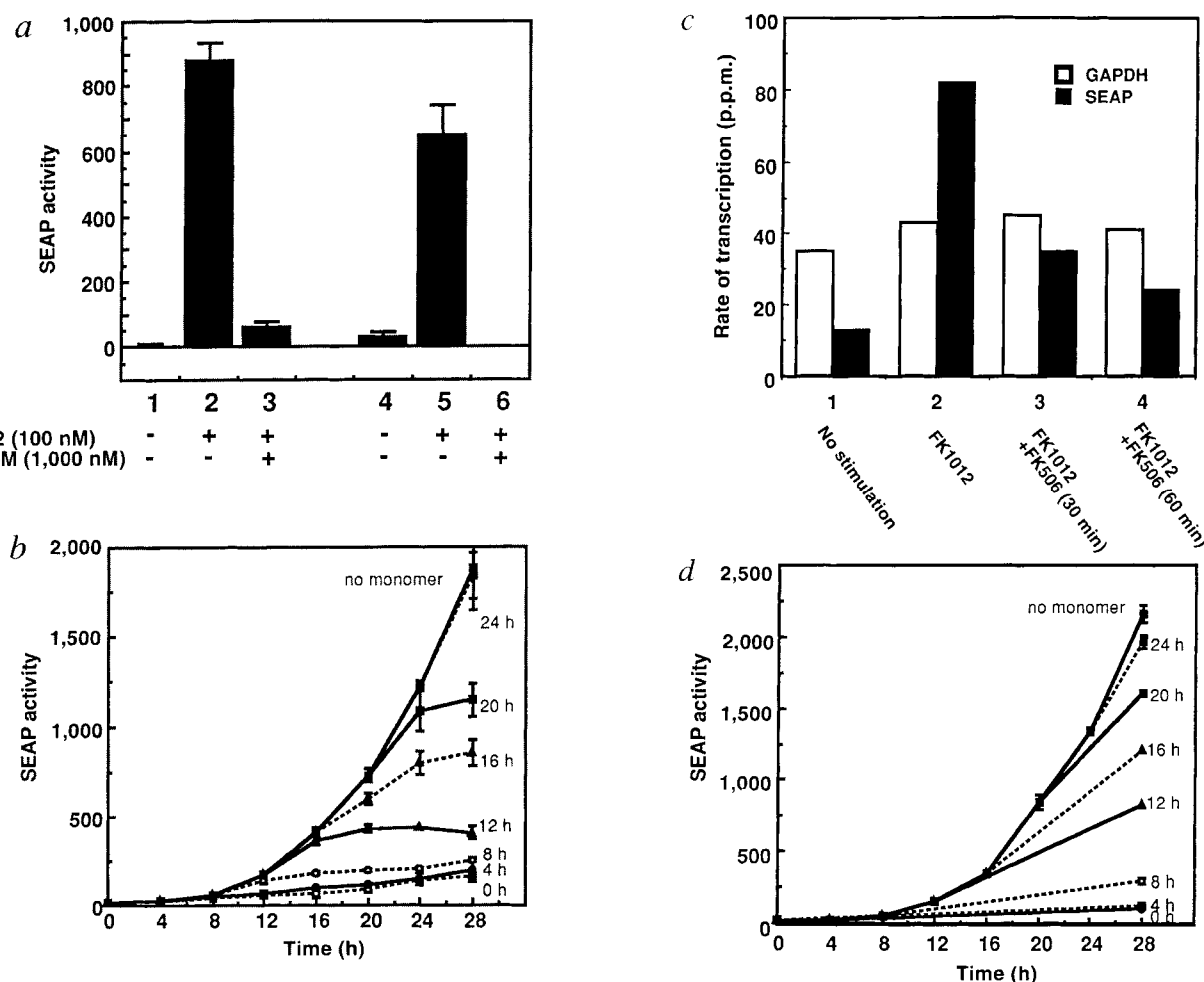


FIG. 2 Transcription induced by a dimeric ligand is effectively competed by monomeric ligand. **a**, Monomeric ligand inhibits transcription induced by dimeric ligand. Cells were transfected with either 1 μ g GF3 plus 10 μ g NF3V1 (lanes 1–3) or 1 μ g HF3 plus 10 μ g NF3V1 (lanes 4–6). Dimeric ligand (FK1012) with or without monomeric ligand (FK506-M) was added as indicated 24 h after transfection. **b**, Monomeric ligand rapidly terminates transcription previously induced by dimeric ligand. Transcription was initiated by addition of FK1012 (10 nM) to Jurkat TAg cells transiently transfected with GF3 (1 μ g), NF3V1 (10 μ g) and G5IL2SX (3 μ g). Excess

FK506-M (100 nM) was added at the time points indicated to the right of each line, and accumulation of SEAP was measured at various time points before and after addition of monomeric ligand. **c**, Monomeric ligand rapidly terminates on-going transcription measured in nuclear run-on assays. Jurkat cells were transfected as above, stimulated with 20 nM FK1012 for 9 h, and 2 μ M FK506 was added for the indicated periods of time. **d**, Monomeric ligand rapidly terminates transcription initiated from an integrated promoter. Transcription was initiated by addition of dimeric ligand to HT1080 clone B cells. Monomeric ligand was added as above.

FIG. 3 Controlling transcription of integrated genes in *Saccharomyces cerevisiae* with heterodimerizing ligands. **a**, Activation of transcription by FK506 depends on the FKBP and calcineurin A fusion proteins. Incubation with FK506 for 3.5 h induces the dose-dependent expression of a *GAL1-lacZ* reporter in cells containing both the GAL4BD-hFKBP12 and GAL4AD-CNA fusions (open squares), but not in cells containing GAL4BD (pAS1) and GAL4AD-CNA (filled diamonds) or GAL4BD-hFKBP12 and GAL4AD (pSE1107) (filled triangles). **b**, Activation of transcription by rapamycin depends on the FKBP and FRB fusion proteins. Incubation with rapamycin induces the dose-dependent expression of a *GAL1-lacZ* reporter in cells containing both the GAL4BD-hFKBP12 and GAL4AD-FRB fusions (open squares), but not in cells containing the GAL4BD-hFKBP12 and GAL4AD-CNA fusions (filled circles). **c**, Rapamycin blocks FK506-induced transcription. Simultaneous addition of increasing amounts of rapamycin (rap, relative excess by weight) inhibits transcription of a *GAL1-lacZ* reporter induced by 50 ng ml⁻¹ FK506 in a 4-h incubation.

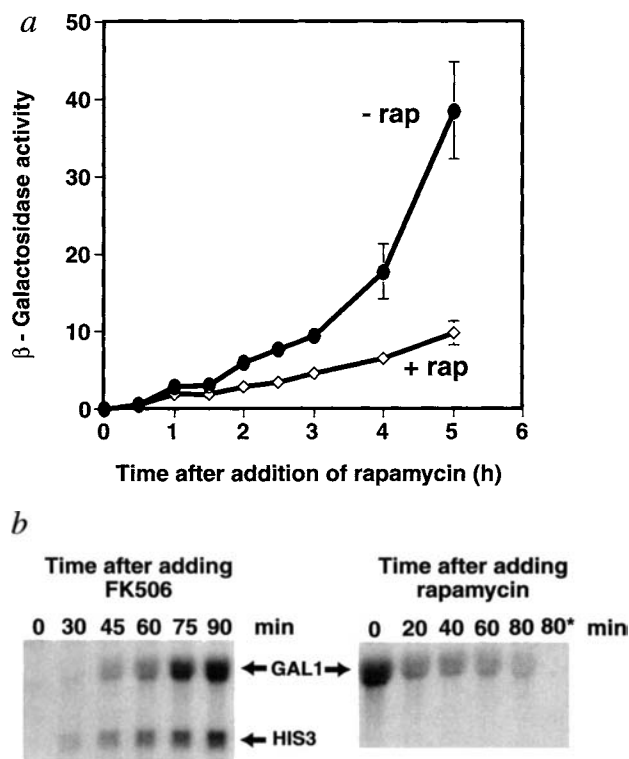
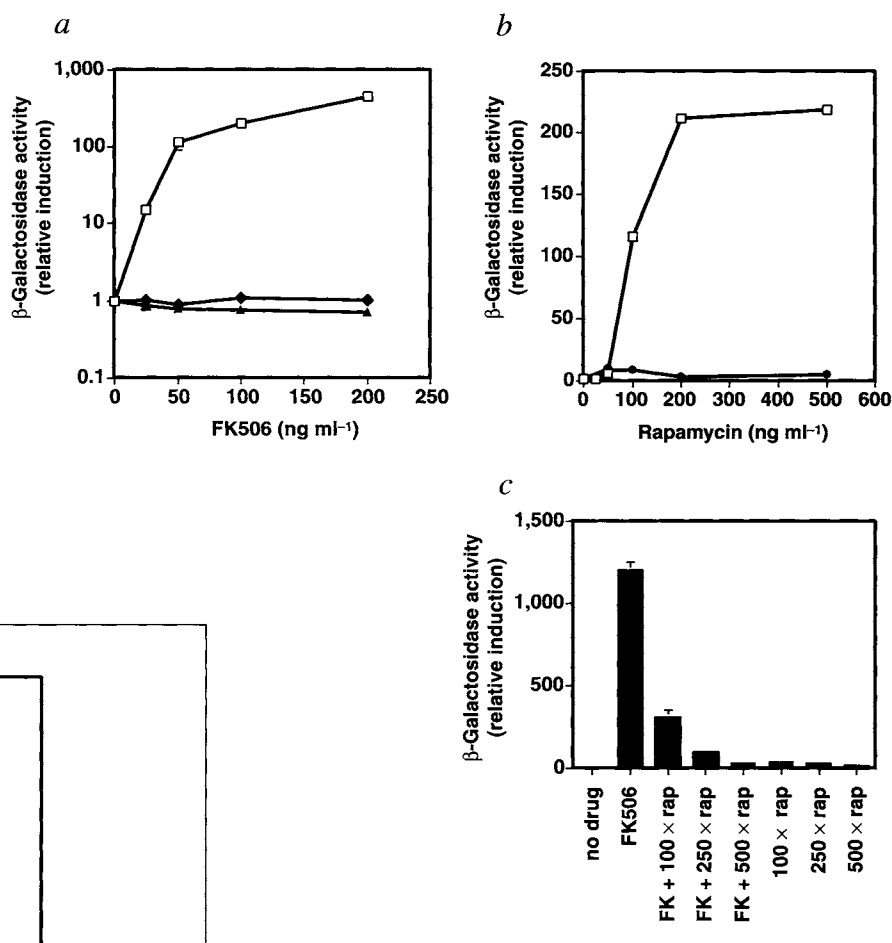


FIG. 4 On-going transcription of integrated and endogenous genes requires the presence of an activation domain. **a**, Accumulation of β -gal activity with and without excess rapamycin. Transcription of the integrated *GAL1-lacZ* reporter gene was induced with 50 ng mg⁻¹ FK506. After 1 h of induction with FK506, a 300-fold excess (by weight) of rapamycin was added to half of the culture, and accumulation of β -galactosidase in the presence (+rap) and absence (–rap) of rapamycin was measured in cells removed from culture at the indicated time points. **b**, Association and dissociation of the activation domain results in rapid activation and termination of transcription from single-copy genes. Induction of transcription of the endogenous *GAL1* and integrated *GAL1-HIS3* genes was measured by northern analysis of RNA from cells collected at the indicated times after addition of 50 ng ml⁻¹ FK506 (left). Excess rapamycin (25 μ g ml⁻¹) was added to cultures induced to express *GAL1* with 150 ng ml⁻¹ FK506 for 14.5 h (right). *GAL1* RNA was measured at the indicated times after addition of rapamycin. The asterisk indicates the level of induction of *GAL1* RNA after treatment with rapamycin alone for 80 min.

Methods

Constructs. Mammalian expression constructs are based on the pBJ5 vector, a derivative of pCD-SR α ²⁶, and contain complementary DNA domains generated using PCR primers that introduced unique restriction endonuclease sites to allow direct subcloning. DNA-binding domain constructs consist of either the GAL4 DNA-binding domain (amino acids 1–147) or the HNF1 DNA-binding domain (amino acids 1–282) linked at the carboxy terminus to three repeated FKBP12 domains (amino acids 2–107), designated GF3 and HF3, respectively. The transcriptional activation domain construct NF3V1 consists of the VP16 trans-activation domain (amino acids 413–490) linked at the amino terminus to three repeated FKBP12 domains with an N-terminal nuclear localization sequence from SV40 large T antigen. The DNA sequence of all constructs was verified by dideoxy sequencing, and full-length protein expression was verified by immunoblotting with the anti-haemagglutinin epitope monoclonal antibody, 12CA5. Yeast expression constructs, including GAL4BD-hFKBP12 and GAL4AD-CNA fusion constructs²⁷, and pAS1 and pSE1107 (ref. 28), have been described previously. The GAL4AD-FRB construct consist of the FKBP-rapamycin binding domain (*M*, 11K) from FRAP¹⁷ joined to the GAL4 activation domain.

Reporter gene analysis. Analysis of the mammalian expression constructs was performed by transient transfection of Jurkat cells stably expressing the SV40 T antigen²⁹. The indicated plasmids were transfected (Biorad Gene Pulser, 250 V, 960 μ F, 0.4 cm cuvette) together with 4 μ g of the SEAP reporter plasmid in which transcription is directed by either 5 GAL4 binding sites (G5IL2SX) or 3 HNF1 binding sites (β 28IL-2SX) within the human interleukin-2 minimal promoter. Total DNA transfected was equalized by addition of vector plasmid. FK1012 or FK506-M (monomeric FK506 with an attached bis-*p*-xylylene carbamate moiety)²³ was added at the indicated times 24 h after transfection, and SEAP was measured approximately 24 h later. The HT1080 clone B-cell line, a human fibrosarcoma line in which G5IL2SX has been stably integrated, was transiently transfected with GF3 (0.5 μ g) and NF3V1 (2 μ g) using lipofectamine (GIBCO BRL). Data shown are representative of at least four independent transfections. The yeast β -galactosidase reporter gene activity was measured by routine methods, with the exception that *ortho*-nitrophenyl galactoside was replaced with 0.3 mM methyl umbelliferyl galactoside (Sigma) and β -galactosidase activity was measured by fluorescence. In Fig. 3, β -galactosidase activity is plotted relative to cultures with no FK506 added, with all measurements being normalized to the absorbance at 600 nm (A_{600}) of the cultures. In Fig. 4a, β -galactosidase activity was normalized to the A_{600} of the cultures and plotted as

induction relative to 0.0002 units of purified β -galactosidase (Sigma). All data are representative of two independent transformants.

Nuclear run-on analysis. Nuclear run-on experiments were performed by transient transfection of the induced template. Nuclei were isolated and used immediately in the run-on reaction to minimize variability associated with a transiently transfected template. Radiolabelled RNA was hybridized to immobilized cDNA fragments (5 μ g each) prepared by PCR amplification. Additional controls not shown include luciferase cDNA, hygromycin B phosphotransferase (HPT) cDNA (Jurkat TAG cells constitutively express HPT²⁹), and a DNA fragment from the promoter of the SEAP gene as a control for read-through and opposite-strand transcription. The rate of transcription of the transfected SEAP gene (determined by subtraction of the rate of transcription of the SEAP promoter DNA fragment from that of the SEAP cDNA fragment, after correcting for the relative length of the DNA fragments) and the endogenous glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene is expressed as amount of radioactivity specifically incorporated relative to total RNA synthesized (p.p.m.). Data shown are representative of three independent experiments.

Yeast strain. The yeast strain YDF6 (*fpr1 Δ ::ADE2 TOR1-1*) (derived from Y153)²⁸ carries a deletion of *FPR1*, thereby minimizing non-productive interactions with endogenous FKBP12, and a dominant *TOR1* allele that renders the cells resistant to rapamycin³⁰. Experiments were performed with mid-log phase cultures carrying the various yeast expression plasmids.

Received 7 June; accepted 26 July 1996.

- Horikoshi, M., Hai, T., Lin, Y. S., Green, M. R. & Roeder, R. G. *Cell* **54**, 1033–1042 (1988).
- Lin, Y. S. & Green, M. R. *Cell* **64**, 971–981 (1991).
- Klein, C. & Struhl, K. *Science* **266**, 280–282 (1994).
- Zawel, L. & Reinberg, D. *Annu. Rev. Biochem.* **64**, 533–561 (1995).
- Stringer, K. F., Ingles, C. J. & Greenblatt, J. *Nature* **345**, 783–786 (1990).

- Lin, Y. S., Hai, T., Maldonado, E., Reinberg, D. & Green, M. R. *Nature* **353**, 569–571 (1991).
- Tjian, R. & Maniatis, T. *Cell* **77**, 5–8 (1994).
- White, J. et al. *EMBO J.* **11**, 2229–2240 (1992).
- Hai, T. W., Horikoshi, M., Roeder, R. G. & Green, M. R. *Cell* **54**, 1043–1051 (1988).
- Brent, R. & Ptashne, M. *Cell* **43**, 729–736 (1985).
- Keegan, L., Gill, G. & Ptashne, M. *Science* **231**, 699–704 (1986).
- Fields, S. & Song, O. *Nature* **340**, 245–246 (1989).
- Spencer, D. M., Wandless, T. J., Schreiber, S. L. & Crabtree, G. R. *Science* **262**, 1019–1024 (1993).
- Liu, J. et al. *Cell* **66**, 807–815 (1991).
- Griffith, J. P. et al. *Cell* **82**, 507–522 (1995).
- Bierer, B. E. et al. *Proc. Natl Acad. Sci. USA* **87**, 9231–9235 (1990).
- Chen, J., Zheng, X. F., Brown, E. J. & Schreiber, S. L. *Proc. Natl Acad. Sci. USA* **92**, 4947–4951 (1995).
- Johnston, M., Flick, J. S. & Pexton, T. *Mol. Cell. Biol.* **14**, 3834–3841 (1994).
- Hawley, D. K. & Roeder, R. G. *J. Biol. Chem.* **262**, 3452–3461 (1987).
- Jiang, Y. & Gralla, J. D. *Mol. Cell. Biol.* **13**, 4572–4577 (1993).
- Szentirmay, M. N. & Sawadogo, M. *EMBO J.* **12**, 4677–4684 (1993).
- Zawel, L., Kumar, K. P. & Reinberg, D. *Genes Dev.* **9**, 1479–1490 (1995).
- Roberts, S. G., Choy, B., Walker, S. S., Lin, Y. S. & Green, M. R. *Curr. Biol.* **5**, 508–516 (1995).
- Chatterjee, S. & Struhl, K. *Nature* **374**, 820–822 (1995).
- Klages, N. & Strubin, M. *Nature* **374**, 822–823 (1995).
- Takebe, Y. et al. *Mol. Cell. Biol.* **8**, 466–472 (1988).
- Cipstone, N. A., Fiorentino, D. F. & Crabtree, G. R. *J. Biol. Chem.* **269**, 26431–26437 (1994).
- Durfee, T. et al. *Genes Dev.* **7**, 555–569 (1993).
- Northrop, J. P., Ullman, K. S. & Crabtree, G. R. *J. Biol. Chem.* **268**, 2917–2923 (1993).
- Cafferkey, R. et al. *Mol. Cell. Biol.* **13**, 6012–6023 (1993).
- Rivera et al. *Nature Medicine* (in the press).

ACKNOWLEDGEMENTS. We thank V. Rivera for providing the HT1080 clone B-cell line. S.N.H. was the recipient of a Walter V. and Idun Y. Berry postdoctoral fellowship. S.R.B. was supported by a medical scientist scholarship from the Life and Health Insurance Medical Research Fund. This work was supported by the Howard Hughes Medical Institute (S.L.S. and G.R.C.).

CORRESPONDENCE and requests for materials should be addressed to G.R.C. (e-mail: hf.grc@forsthe.stanford.edu). All reagents, including FK1012, will be made available on request.

Positional cloning of the mouse retrovirus restriction gene *Fv1*

Steve Best*, Paul Le Tissier, Greg Towers & Jonathan P. Stoye

Division of Virology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA UK

VERTEBRATE evolution has taken place against a background of constant retrovirus infection, and much of the mammalian genome consists of endogenous retrovirus-like elements¹. Several host genes have evolved to control retrovirus replication², including *Friend-virus-susceptibility-1*, *Fv1*, on mouse chromosome 4 (refs 3, 4). The *Fv1* gene acts on murine leukaemia virus at a stage after entry into the target cell but before integration and formation of the provirus⁵. Although restriction is not absolute, *Fv1* prevents or delays spontaneous or experimentally induced viral tumours². *In vitro*, *Fv1* restriction leads to an apparent 50–1,000 fold reduction in viral titre⁶. Genetic evidence implicates a direct interaction between the *Fv1* gene product and a component of the viral preintegration complex, the capsid protein CA (refs 7–9). We have now cloned *Fv1*: the gene appears to be derived from the *gag* region of an endogenous retrovirus unrelated to murine leukaemia virus, implying that the *Fv1* protein and its target may share functional similarities despite the absence of nucleotide-sequence homology.

The genetic properties of *Fv1* lend themselves to a positional cloning strategy. *Fv1* and closely linked genes have been accurately mapped^{10,11}; multiple alleles of *Fv1* and viral determinants allow independent phenotypic tests. Most inbred strains carry either the *Fv1ⁿ* or the *Fv1^b* allele. Mice of the *Fv1ⁿ* genotype are susceptible to so-called N-tropic murine leukaemia viruses (MLVs) and resistant to B-tropic MLVs¹² and vice versa. NB-tropic viruses are not restricted by either allele. Moreover, *Fv1* is a dominant gene³, allowing functional assays in cell lines transfected

with cloned DNAs such as yeast artificial chromosomes (YACs).

Candidate YACs were identified using two markers, the endogenous provirus, *Xmv9*, and the gene *Nppa*, believed to lie within 1.2 megabases (Mb) of *Fv1* (refs 10, 11). These YACs were introduced into mouse L cells and tested for the presence of *Fv1* (Fig. 1). Cells were infected with N- and B-tropic viruses carrying resistance to puromycin with virus stocks which gave equal numbers of puromycin-resistant colonies on *Fv1*-null cells derived from *Mus dunni*¹³. As expected, given that L cells are derived from an *Fv1ⁿ* strain and will restrict B-tropic MLV, both control and *Xmv9* J1 cells show about tenfold fewer colonies with the B-tropic virus than the N-tropic virus. By contrast, cells carrying *Nppa* D11 showed low colony numbers with both N and B viruses. Six independent G418-resistant lines derived by fusion with *Nppa* D11 were tested; of these, five showed the same reduction in titre of the N-tropic virus with no effect on an NB-tropic control (not shown).

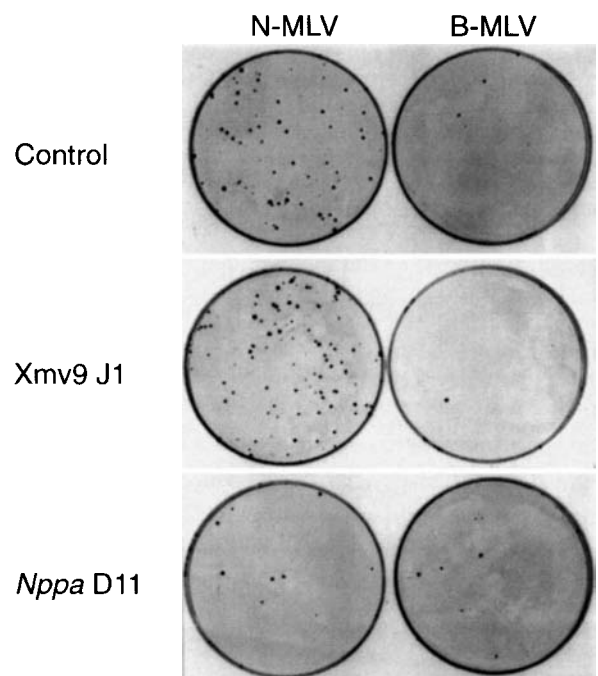
A cosmid library was prepared from YAC *Nppa* D11. Twenty randomly selected cosmids were introduced into cells and clones from each transfection tested for *Fv1* activity. Three overlapping cosmids, D11C5, D11C11 and D11C44, were identified which carried activity (Fig. 2a). Non-overlapping cosmids lacked *Fv1* activity. An extensive series of negative clones obtained with cosmid D11C63 suggested that *Fv1* must lie on the right-hand end of D11C5. Analysis of two smaller constructs derived from this region of D11C5 showed that the complete *Fv1* gene must lie within a 6.5-kb *SpeI*–*EcoRI* fragment. Sequence analysis revealed an open reading frame (ORF) of ~1,400 nucleotides within this fragment. A methionine, within an acceptable Kozak consensus, was present at the seventh codon of the ORF. The ORF was cloned into an expression vector and transfected into *M. dunni* cells. Five of seven G418-resistant cell clones restricted N-MLV but failed to restrict NB- or B-tropic MLV (Fig. 2b), confirming that we had cloned the *b* allele of *Fv1*. This analysis indicated that the *Fv1* gene comprised a single exon encoding a protein product 459 amino acids in length with a predicted relative molecular mass of 52K (Fig. 3a).

A 9-kb *EcoRI* fragment corresponding to the *Fv1ⁿ* allele was isolated from an AKR genomic library. Restriction analysis and sequencing showed two gross changes in the DNA structure

* Present address: University of Oxford, The Wellcome Trust Centre for Human Genetics, Windmill Road, Headington, Oxford OX3 7BN, UK.

FIG. 1 Identification of a YAC with *Fv1* activity. Colonies of puromycin-resistant control L cells or cells transformed with the Xmv9J1 and Nppa D11 YACs infected with equal titres of N- or B-tropic MLV vectors carrying the puromycin-resistance gene.

METHODS. YAC libraries^{20,21} were screened for *Xmv9* (primers SB1 and SB3; ref. 11) and *Nppa* (primers 5'-GAACCTGCTAGACGACCTGG-3' and 5'-AAGCTGTTCAGCAGCTAGTCC-3') by PCR as described²⁰. Three YACs, Nppa G7 (210 kb), D11 (190 kb) and Xmv9 B2 (250 kb) from the Princeton library²⁰ and one, Xmv9J1 (350 kb) from the St Mary's library²¹, were chosen for further study. The YAC Xmv9J1 was transferred to a RAD52 background by mating with a non-YAC-containing RAD52 strain followed by sporulation and tetrad dissection. YACs were then modified to carry the *neo'* gene by homologous recombination with linearized pRV1.²² Yeast spheroplasts were fused with mouse cells essentially as described²³ and colonies of neomycin-resistant cells isolated after 2 weeks growth in 800 $\mu\text{g ml}^{-1}$ G418. L cells yielded 20 times more colonies than fibroblastoid lines and also took up a significantly higher fraction of potentially intact YACs, as judged by PCR assays for the YAC arms and internal marker. L cells spontaneously release endogenous ecotropic virus (not shown) which interferes with infection by endogenous ecotropic MLV²⁴ but does not affect *Fv1* function²⁵. Assays for *Fv1* therefore utilized virus with polytropic envelope gene, which does not cross-interfere with ecotropic virus²⁴, to deliver a vector carrying resistance to puromycin. Polytropic N- and B-tropic MLVs were constructed *in vitro* by replacing a 2.9-kb *Pst*I (in the LTR)-*i*-HindIII (in *pol*) fragment from p247-W²⁶ with the analogous *Pst*I-HindIII fragments from the ecotropic N- (pWN41) and B-tropic (pWB5) MLV clones²⁷. Virus stocks conferring resistance to puromycin were prepared by lipofectin-mediated, cotransfection of *M. dunni* cells¹³ with pBABEpuro²⁸ together with cloned viral DNA, followed by selection for 14 days in puromycin (5 $\mu\text{g ml}^{-1}$), infection of fresh *M. dunni* cells and selection of reverse-transcriptase-positive²⁹, puromycin-resistant colonies. Virus stocks were titrated on *Fv1*-null 3T3FL or *M. dunni* cells as well as n-type NIH-3T3 cells and b-type BALB-3T3 cells. Virus infections used 15 $\mu\text{g ml}^{-1}$ poly-



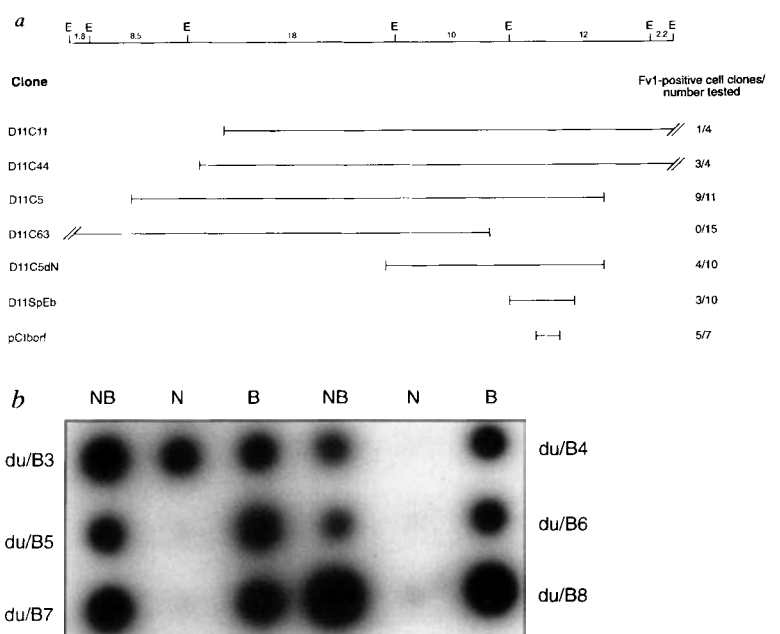
brene. Stocks showed appropriate relative sensitivities to *Fv1* restriction in this assay, with restriction manifested by a 10–20-fold reduction in titre.

between the analogous fragments from AKR and C57BL/6 (Fig. 3b): (1) insertion of an intracisternal A-type particle (IAP) element about 1.8 kb downstream from the ORF within a pair of B2 repeats and not expected to affect the protein product of the putative gene; and (2) a 1.3-kb deletion of DNA in AKR compared to C57BL/6. This removes the final 22 amino acids of the

predicted protein, substituting in their place three different amino acids (Fig. 3c), implying that the *n* allele of *Fv1* encodes a protein 19 amino acids shorter than the *b* allele. *Fv1* from C3H mice, a second *Fv1*ⁿ strain, had the same changes as AKR compared to C57BL/6 and BALB/c, two *Fv1*^b strains. Two other predicted single-amino-acid changes were also seen in the AKR ORF that

FIG. 2 Sub-cloning the *Fv1* gene. *a*, Overlapping cosmid clones and derivatives originating from YAC Nppa D11 were mapped and tested for *Fv1* activity. The ORF with *Fv1* activity runs from right to left. Numbers are kilobases. E, *Eco*RI sites. *b*, *Fv1* assays for six independent cell lines (du/B3–du/B8) transfected with pCIBorf.

METHODS. A cosmid library from the C57BL/6/J YAC D11 was prepared. Blocks of total yeast DNA containing YAC D11 were treated with increasing concentrations of *Mbo*I and DNA from a portion of each block was sized by pulsed-field electrophoresis. DNA was extracted from the remaining portion of blocks that contained DNA of appropriate size and cloned into SuperCos1 (Stratagene). The cosmid library was screened using random-primed total mouse DNA. Overlapping cosmids were identified by side-by-side comparison of *Eco*RI- and *Bgl*II-digested cosmid DNAs. Two subclones were derived from cosmid D11C5. D11C5dN was obtained after partial *Not*I digestion of a derivative of the starting cosmid into which a *Not*I site had been introduced via transposon mutagenesis³⁰. The second, D11SpEb, represents a 6.5-kb *Spe*I to *Eco*RI fragment from D11C5 cloned into pBK-CMV (Stratagene). The *Fv1* ORF was PCR-amplified from D11SpEb using *Pfu* polymerase in combination with primers spanning the 5' (5'-GACGCGCTAGCCGAGTCTAGG-GAAA-3') and 3' (5'-GGATATGTCGACTCCTCCTCTGATTTAA-3') ends of the predicted open reading frame, cloned into pCR-Script (Stratagene) at the *Srf*I site and transferred to pCI-neo (Promega) following *Eco*RI plus *Not*I digestion, to yield pCIBorf. Sequence analysis showed that D11C5 and pCIBorf contained identical ORFs. *Fv1*-containing clones were introduced into NIH3T3 cells (*a*) or *M. dunni* (*b*), and individual G418-resistant cell clones were selected and tested for *Fv1* by infection with N, B and NB-tropic MLVs, followed 3 days later by reverse transcriptase



assays. Note that not all cell clones derived from by transformation with a plasmid containing *Fv1* showed restriction; expression of the phenotype was strictly correlated with transcription of the intact ORF.

resulted in changes at amino acids 358 and 399. Expression of the AKR ORF in *M. dunnii* cells resulted in the inhibition of B-tropic MLV but did not affect replication of NB- or N-tropic MLV (data not shown), the predicted result for the *Fv1ⁿ* allele.

We also examined the *Fv1* gene from the DBA/2 mouse strain that carries the *Fv1ⁿ* allele but gives slightly different titration curves for *Fv1* inhibition¹⁴. DBA/2 turned out to have the same predicted protein product as AKR but differed in non-coding sequences, lacking the IAP element found in AKR (Fig. 3b). The presence or absence of this element may be responsible for the

subtle difference in phenotype between DBA/2 and other *Fv1ⁿ* strains, perhaps affecting messenger RNA stability.

Although we readily detected transcription of the *Fv1* locus by using reverse transcription with the polymerase chain reaction (RT-PCR), we could not detect a product by northern blotting, suggesting that levels of gene expression were low. This complicates the mapping of the 5' and 3' ends of the *Fv1* transcript. There are several potential Sp1-binding sites in the 800 nucleotides preceding the ORF, and the GeneID program¹⁵ predicts two Pol II promoter sites in this region. But no TATA box is present in

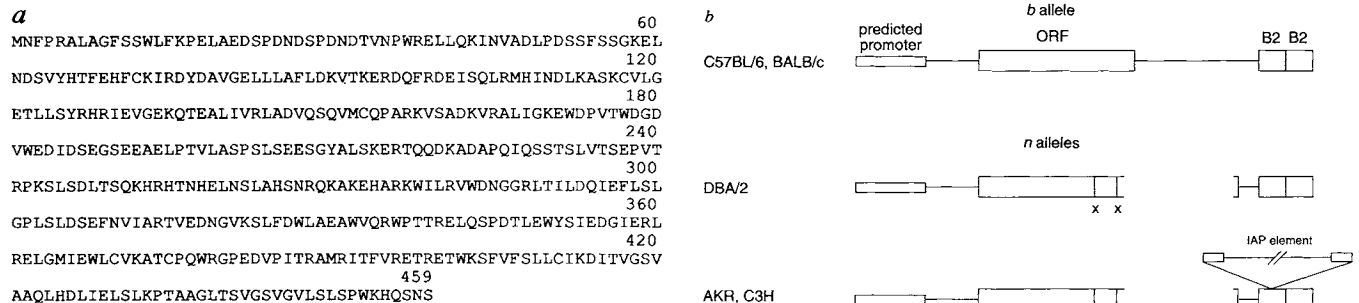


FIG. 3 Structure of the *Fv1* gene. **a**, Predicted structure of the *Fv1* ORF (single-letter amino-acid code). The sequence of a 6.48-kb *SpeI*-*EcoRI* fragment from cosmid D11C5 was determined (EMBL accession number, X97719) and an ORF running from nucleotides 2,149 to 3,543 of this fragment was identified. The seventh codon of the ORF encodes the methionine at position 1. **b**, Structure of the *Fv1* gene in C57BL/6, BALB/c, DBA/2, C3H and AKR mice. The positions of the predicted promoter, ORF and insertion elements providing polyadenylation sites are shown. Single amino-acid differences are marked. X; *n* alleles have a 1.3-kb deletion toward the 3' end of the ORF, resulting in an altered C terminus. **c**, Predicted amino-acid differences between the *Fv1* genes in C57BL/6 and AKR.

METHODS. DNA sequences were determined by the dideoxy method with Sequenase 2.0 from M13 subclones or single-stranded phage DNA rescued from f1-origin containing plasmids using helper phage M13 K07. The AKR/J *Fv1* clone was isolated from an *EcoRI* library prepared in the lambda vector DASH II (Stratagene). The library was screened using a 190-bp fragment (synthesized by PCR using the primers 5'-TTAGCCTG-GAGCTCCTGAAG-3' and 5'-GAAGCGGAAGAAGTCTCTTG-3') which maps to the extreme right-hand end of D11C5. A region extending through the *Fv1*

ORF into the IAP LTR was sequenced (EMBL accession number, X97720). The DBA/2 allele was synthesized by long PCR using primers 5'-GAAGCG-GAAGAAGTCTCTTG-3' plus 5'-CATTTGCTCATCTGTGCCAG-3' and has also been sequenced. The BALB/c and C3H alleles were synthesized by long PCR and examined by restriction mapping. The activity of the AKR *Fv1* ORF was assayed as for *Fv1^b* using a subclone in pCI-neo.

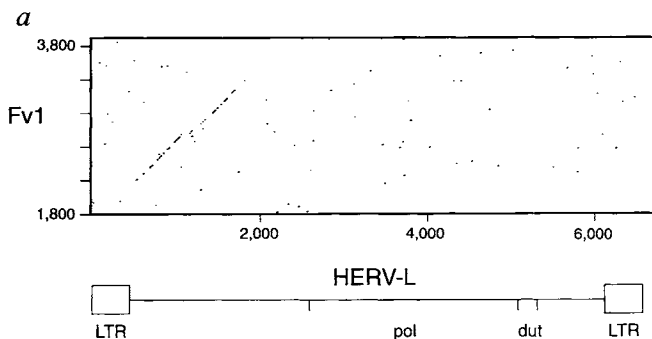


FIG. 4 Origin of *Fv1*. **a**, Dot-matrix comparison between *Fv1* and HERV-L using the COMPARE program of the GCG package. No homology was seen between the sequences up- and downstream of the *Fv1* ORF (nucleotides 2,149–3,543) and HERV-L. Only the polymerase (pol) and dUTPase (dut) genes of HERV-L show homology to other retroviruses¹⁶. **b**, Cross-species hybridization of *Fv1*. Genomic DNA from mouse (1, C57BL/6; 2, AKR; 3, *M. dunnii*), human, rat and cat was digested with *HindIII* (left 7 lanes) or *EcoRI* (right 7 lanes), fractionated on an agarose gel, transferred to nitrocellulose and probed with the *Fv1* ORF. Hybridization was at 65 °C, washes were in 4 × SSC.0.1%SDS at 65 °C. Smeared signals are present in lanes containing the human and mouse, and to a lesser extent the rat, samples; these smears disappeared on more stringent washing (0.2 × SSC/0.1%SDS at 65 °C). The *EcoRI*-digested *M. dunnii* sample was underloaded.

this part of the genomic sequence and we have not identified a unique primer extension or 5' rapid amplification of cloned ends (RACE) product. However, we have cloned the presumptive 3' ends of the *n* and *b* transcripts using 3' RACE. The *Fv1*ⁿ allele of AKR terminates at the R-U5 boundary of the 5' long terminal repeat (LTR) of the inserted IAP, whereas the *b* allele is polyadenylated within the second B2 element downstream of the ORF. The *Fv1* mRNA is unusual, therefore, in having a single large exon, encoding a protein of ~50K and transcripts terminating within insertion elements found downstream of the ORF.

Sequence homology was detected with two expressed sequence tags (accession numbers X71645 and R99971), as well as a genomic clone (X89211) representing a member of the human endogenous retrovirus family, HERV-L¹⁶ (Fig. 4a). *Fv1* and HERV-L show 60% identity in the 5' region of the presumptive *gag* gene of HERV-L. Hybridization between *Fv1* and the HERV-L provirus, which is present at 200 copies per haploid genome equivalent in human and mouse, but at a much lower level in rat¹⁶, may explain the background smear seen in low-stringency Southern blots. Thus, *Fv1* is likely to be derived from a retrovirus of the HERV-L family.

Fv1 and the endogenous human retrovirus HERV-L are 60% identical over a stretch of 1,300 base pairs. This region of the cloned HERV-L does not encode an ORF, presumably reflecting mutational decay of the endogenous retrovirus, and shows no discernible sequence homology to other retroviruses. But based on its position between an LTR and a *pol* gene, it seems that it corresponds to the *gag* gene of the HERV-L family. The *Fv1* protein may therefore be Gag-like and share some structural similarity with its target, the CA protein of MLV. Although this presumed similarity in *gag* might be important for virus restriction, for example to permit binding of *Fv1* to CA, there are hundreds of retroviral elements more closely related to MLV than *Fv1* and none of these restricts replication.

To study the evolution of *Fv1* further we examined DNAs from a variety of mammals. Single-copy bands were seen in samples of inbred mice and in *Mus dunni* (Fig. 4b), but no reactive bands were seen in samples from rat, cat and human. The *Fv1* gene is therefore present only in mice, or it shows a high degree of sequence divergence even between mouse and rat. The overall similarity of the *Fv1* locus in *Mus dunni* (not shown) and *Mus musculus* (C57BL/6) mice implies that the original integration event preceded the divergence of the progenitors of the two species and that the *Fv1* gene is likely to be widely distributed among *Mus* species. Endogenous MLVs are less widely distributed¹⁷; they appear to have colonized the *Mus* germ line more recently. *Fv1* activity is found only in strains containing endogenous MLVs¹⁸, suggesting that the antiviral activity may have evolved in response to the presence of MLVs. The apparently still-more-recent generation of the *Fv1*ⁿ allele suggests that this may be an on-going process. Given the widespread distribution of retroviral elements, it is perhaps not surprising that genes from these elements can be co-opted for the purposes of the host, for example in protecting against retroviral infection¹⁹. The demonstration of such activities in mice raises the question of whether analogous restriction systems have evolved elsewhere, for instance in humans, where the absence of inbred strains would have obscured their presence. □

Received 14 May; accepted 8 July 1996.

1. Baltimore, D. *Cell* **40**, 481–482 (1985).
2. Steeves, R. & Lilly, F. *Annu. Rev. Genet.* **11**, 277–296 (1977).
3. Lilly, F. *J. Natl Cancer Inst.* **45**, 163–169 (1970).
4. Rowe, W. P., Humphrey, J. B. & Lilly, F. *J. Exp. Med.* **137**, 850–853 (1973).
5. Pryciak, P. M. & Varmus, H. E. *J. Virol.* **66**, 5959–5966 (1992).
6. Hartley, J. W., Rowe, W. P. & Huebner, R. J. *J. Virol.* **5**, 221–225 (1970).
7. Boone, L. R., Innes, C. L., Glover, P. L. & Linney, E. J. *J. Virol.* **63**, 2592–2597 (1989).
8. Bowerman, B., Brown, P. O., Bishop, J. M. & Varmus, H. E. *Genes Dev.* **3**, 469–478 (1989).
9. Rommelaere, J., Donis-Keller, H. & Hopkins, N. *Cell* **16**, 43–50 (1979).
10. Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. *J. Virol.* **63**, 1763–1774 (1989).
11. Stoye, J. P., Kaushik, N., Jeremiah, S. & Best, S. *Mammalian Genome* **6**, 31–36 (1995).
12. Jolicoeur, P. *Curr. Top. Microbiol. Immunol.* **86**, 67–122 (1979).
13. Lander, M. R. & Chattopadhyay, S. K. *J. Virol.* **52**, 695–698 (1984).

14. Pincus, T., Hartley, J. W. & Rowe, W. P. *Virology* **65**, 333–342 (1975).
15. Guigo, R., Knudsen, S., Drake, N. & Smith, T. J. *Mol. Biol.* **226**, 141–157 (1992).
16. Cordonnier, A., Casella, J.-F. & Heidmann, T. *J. Virol.* **69**, 5890–5897 (1995).
17. Kozak, C. A. & O'Neill, R. R. *J. Virol.* **61**, 3082–3088 (1987).
18. Kozak, C. A. *J. Virol.* **55**, 281–285 (1985).
19. Gardner, M. B., Kozak, C. A. & O'Brien, S. J. *Trends Genet.* **7**, 22–27 (1991).
20. Burke, D. T., Rossi, J. M., Koos, D. S. & Tilghman, S. M. *Mammalian Genome* **1**, 65 (1991).
21. Chartier, F. L. et al. *Nature Genet.* **1**, 132–136 (1992).
22. Srivastava, A. K. & Schlessinger, D. *Gene* **103**, 53–59 (1991).
23. Pachnis, V., Pevny, L., Rothstein, R. & Constantini, F. *Proc. Natl Acad. Sci. USA* **87**, 5109–5113 (1990).
24. Rein, A. *Virology* **120**, 251–257 (1982).
25. Duran-Troise, G., Bassin, R. H., Wallace, B. F. & Rein, A. *Virology* **112**, 795–799 (1981).
26. Holland, C. A., Wozney, J., Chatis, P. A., Hopkins, N. & Hartley, J. W. *J. Virol.* **53**, 152–157 (1985).
27. Boone, L. R. et al. *J. Virol.* **48**, 110–119 (1983).
28. Morgenstern, J. P. & Land, H. *Nucleic Acids Res.* **18**, 3587–3596 (1990).
29. Goff, S., Traktman, P. & Baltimore, D. *J. Virol.* **38**, 239–248 (1981).
30. Morgan, B. A. et al. *Proc. Natl Acad. Sci. USA* **93**, 2801–2806 (1996).

ACKNOWLEDGEMENTS. This work is dedicated to Frank Lilly. We thank all our colleagues at NIMR, as well as those in the retrovirus, genetics and genomics fields for their help and encouragement; S. Brown, F. Chartier, D. Koos and S. Tilghman for YACs; C. Holland, L. Boone, H. Land and A. Srivastava for plasmids; and J. Coffin, E. Green and M. Rubock for advice. This work was supported by the UK MRC.

CORRESPONDENCE and requests for materials should be addressed to J.P.S. (e-mail: j-stoye@nimr.mrc.ac.uk).

The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry

Conrad C. Bleul*, Michael Farzan†, Hyeryun Choe†, Cristina Parolin†‡, Ian Clark-Lewis§, Joseph Sodroski†|| & Timothy A. Springer*

* The Center for Blood Research and † Division of Human Retrovirology, Dana-Farber Cancer Institute, Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

|| Department of Cancer Biology, Harvard School of Public Health, Boston, Massachusetts 02115, USA

‡ Institute of Microbiology, University of Padua, Padua 35121, Italy
§ Biomedical Research Centre, University of British Columbia, 2222 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada

CHEMOKINES are chemotactic cytokines that activate and direct the migration of leukocytes^{1,2}. There are two subfamilies, the CXC and the CC chemokines. We recently found that the CXC-chemokine stromal cell-derived factor-1 (SDF-1)^{3,4} is a highly efficacious lymphocyte chemoattractant⁵. Chemokines act on responsive leukocyte subsets through G-protein-coupled seven-transmembrane receptors¹, which are also used by distinct strains of HIV-1 as cofactors for viral entry. Laboratory-adapted and some T-cell-line-tropic (T-tropic) primary viruses use the orphan chemokine receptor LESTR/fusin (also known as fusin)^{6–8}, whereas macrophage-tropic primary HIV-1 isolates use CCR-5 and CCR-3 (refs 7–11), which are receptors for known CC chemokines. Testing of potential receptors demonstrated that SDF-1 signalled through, and hence 'adopted', the orphan receptor LESTR, which we therefore designate CXC-chemokine receptor-4 (CXCR-4). SDF-1 induced an increase in intracellular free Ca²⁺ and chemotaxis in CXCR-4-transfected cells. Because SDF-1 is a biological ligand for the HIV-1 entry cofactor LESTR, we tested whether it inhibited HIV-1. SDF-1 inhibited infection by T-tropic HIV-1 of HeLa-CD4 cells, CXCR-4 transfectants, and peripheral blood mononuclear cells (PBMCs), but did not affect CCR-5-mediated infection by macrophage-tropic (M-tropic) and dual-tropic primary HIV-1.

The chemokine SDF-1 has multiple biological activities and may be a primordial chemokine. It was initially defined by cloning as a bone-marrow stromal cell-derived factor, and as a pre-B-cell stimulatory factor^{3,4}. We recently isolated SDF-1 as a T-lymphocyte chemoattractant, and found it to be active on monocytes but not neutrophils⁵. Up to 80% of resting T lymphocytes respond to SDF-1, as opposed to 1–10% for other chemokines, suggesting that its receptor is either highly active or well expressed on resting T lymphocytes. All other known CXC- and CC-chemokines cluster on chromosomes 4 and 17, respectively, but the SDF-1 gene is on chromosome 10 (ref. 12). Sequence homologies show that SDF-1 is related almost equally to both CC- and CXC-chemokines forming an outgroup, and is predicted to have diverged little from a primordial CC- and CXC-chemokine ancestor. Consistent with this, SDF-1 is unusually well conserved between species, with a single substitution of Ile to Val between mouse and human¹². SDF-1 α and SDF-1 β arise by differential splicing from a single gene, and differ in four carboxy-terminal amino acids that are present in SDF-1 β but absent in SDF-1 α .

We isolated two forms of murine SDF-1 from stromal cells that appear to be processed relative to SDF-1 α , as predicted from the cDNA, one of which lacks the five N-terminal amino-acid residues⁵. We synthesized and refolded the corresponding human proteins, the longer SDF-1 α (residues 1–67) and the shorter SDF-1 α (6–67). To test for biological activity of the synthetic proteins, we performed chemotaxis assays^{5,13} with freshly prepared human peripheral blood lymphocytes using Transwell inserts. This allows the exact quantification of migrated versus non-migrated cells^{5,13}, by counting the cells that transmigrate through the filter and drop into the bottom chamber. Full-length SDF-1 α (1–67) induced maximal lymphocyte chemotaxis at a concentration of 1 $\mu\text{g ml}^{-1}$ and was identical in activity to purified natural murine SDF-1 α (Fig. 1). In contrast, SDF-1 α (6–67) was inactive up to a concentration of 3.3 $\mu\text{g ml}^{-1}$. This shows that the N-terminal amino acids are important for SDF-1 function, as has previously been shown for other chemokines¹.

To identify a receptor that binds SDF-1, we screened chemokine receptor-like orphans for their ability to elicit increases in intracellular Ca^{2+} in response to SDF-1. CHO cells stably transfected with human LESTR^{14–18} consistently showed increases in intracellular Ca^{2+} when full-length SDF-1 α (1–67) was added. Ca^{2+} fluxes were seen with SDF-1 concentrations ranging from 1 $\mu\text{g ml}^{-1}$ (Fig. 2a) to 100 ng ml^{-1} (data not shown). In contrast, untransfected cells were unresponsive to SDF-1 (Fig. 2b). Addition of SDF-1 α (1–67) to LESTR-transfectants abrogated responsiveness to a subsequent addition of SDF-1 α (1–67), demonstrating homologous desensitization (Fig. 2a). Consistent with earlier negative functional data¹⁸, addition of RANTES ('regulated on activation normal T cell expressed and secreted'),

monocyte chemoattractant protein (MCP)-1 and interleukin (IL)-8 to a final concentration of 1 $\mu\text{g ml}^{-1}$ did not induce a response. These chemokines also did not desensitize cells to SDF-1 α (1–67) (Fig. 2c). LESTR-transfected CHO cells were unresponsive to addition of N-terminally truncated SDF-1 α (6–67) (Fig. 2d). Furthermore, SDF-1 α (6–67) did not antagonize a response to subsequent stimulation with full-length SDF-1 α (1–67), suggesting that the truncated protein either does not bind LESTR or has a substantially lower affinity than SDF-1 α (1–67). The Ca^{2+} flux assay was confirmed with a modified Boyden-chamber chemotaxis assay. LESTR-transfected CHO cells migrated specifically in response to SDF-1 α (1–67) (Fig. 2e), whereas untransfected cells did not show migration above background (data not shown).

Chemokines and their receptors have recently been shown to play critical roles in early events in HIV-1 infection. The chemokine receptor-like orphan LESTR was identified as a necessary cofactor for entry of T-tropic laboratory-adapted⁶ and T-tropic primary^{7,8} HIV-1 strains into CD4^{+} cells. Having identified SDF-1 as a biological ligand for LESTR, we hypothesized that it would interfere with use of this receptor by HIV-1. To test for inhibition

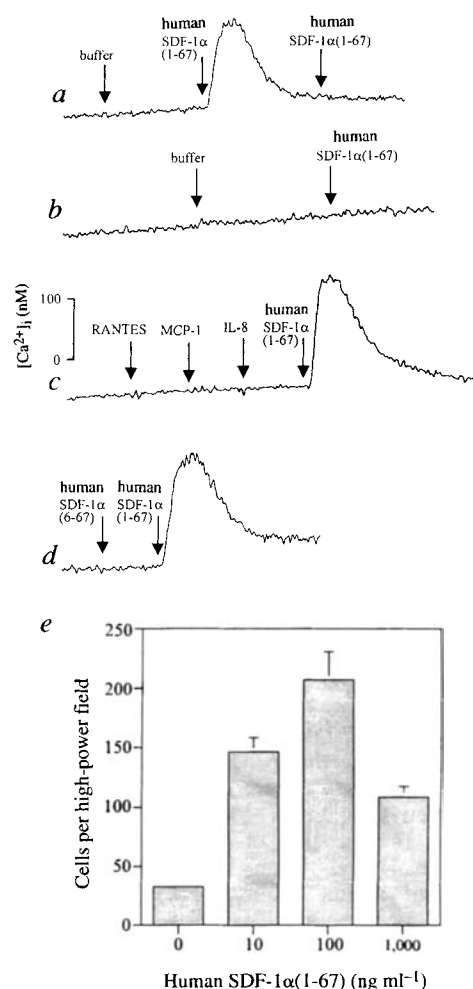


FIG. 2 Mobilization of Ca^{2+} and chemotaxis of LESTR-transfected CHO cells in response to full-length SDF-1 α (1–67) and truncated SDF-1 α (6–67). a–d, A stable LESTR-transfected CHO cell line (a, c, d) and untransfected CHO cells (b) were loaded with Fura-2 AM fluorescent dye, and increases in intracellular Ca^{2+} on addition of the indicated chemokines were measured in a fluorescence spectrophotometer. SDF-1 α (1–67), SDF-1 α (6–67), RANTES, MCP-1 and IL-8 were added to a final concentration of 1 $\mu\text{g ml}^{-1}$. e, Chemotactic activity of SDF-1 α (1–67) in a modified Boyden-chamber chemotaxis assay for a LESTR-transfected CHO cell line. Error bars indicate s.d. of quadruplicates.

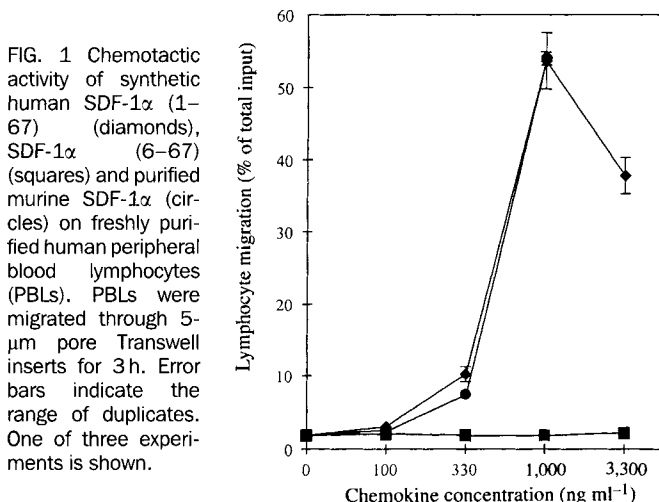


FIG. 1 Chemotactic activity of synthetic human SDF-1 α (1–67) (diamonds), SDF-1 α (6–67) (squares) and purified murine SDF-1 α (circles) on freshly purified human peripheral blood lymphocytes (PBLs). PBLs were migrated through 5- μm pore Transwell inserts for 3 h. Error bars indicate the range of duplicates. One of three experiments is shown.

by SDF-1 of entry of T-tropic HIV-1 strains into permissive cells, we used an *env*-complementation assay^{7,19}, in which HIV-1 envelope glycoproteins expressed in *trans* complement a single round of replication of an *env*-deleted provirus encoding the chloramphenicol acetyltransferase (CAT) gene. We produced recombinant virus that was pseudotyped with the envelope glycoproteins of the laboratory-adapted HXBc2 isolate, which uses LESTR for fusion⁷. Because HeLa cells are known to express LESTR at high levels⁶, we incubated HeLa-CD4 cells, which express human CD4 in the presence of either SDF-1 α (1–67) or SDF-1 α (6–67), with recombinant viruses containing the HXBc2 envelope glycoproteins. The efficiency of the early phase of virus infection was assessed by measuring CAT activity in the target cells 60 h after infection. Full-length SDF-1 α (1–67) completely inhibited infection of the recombinant viruses in a concentration-dependent manner, whereas the truncated SDF-1 α (6–67) showed no inhibi-

tion (Fig. 3a).

To demonstrate that the HIV-1-inhibitory effect of SDF-1 was mediated through interaction with LESTR, we used Cf2Th canine thymocytes as target cells. Cf2Th cells are non-permissive for HIV-1 infection unless transfected with plasmids expressing human CD4 and chemokine receptors specific for particular viral strains⁷. Cf2Th canine thymocytes expressing human CD4 and LESTR were incubated with recombinant virus containing HXBc2 envelope glycoproteins in the presence of either SDF-1 α (1–67) or SDF-1 α (6–67). SDF-1 α (1–67) effectively inhibited infection, whereas truncated SDF-1 α (6–67) had no effect (Fig. 3b). Furthermore, SDF-1 α (1–67) did not inhibit infection of Cf2Th cells expressing CD4 and CCR-5 by a recombinant HIV-1 virus containing the envelope glycoproteins of the M-tropic primary HIV-1 strain ADA (Fig. 3b). These data suggest that functional SDF-1 effectively inhibits entry of HIV-1 virus using LESTR without interfering with infection by HIV-1 strains using CCR-5.

To exclude the possibility that SDF-1 mediated its virus-inhibitory effect by interacting with the viral envelope glycoproteins, we tested recombinant HIV-1 pseudotyped with the envelope glycoproteins of the dual-tropic primary HIV-1 strain 89.6, which is known to use the chemokine receptors LESTR, CCR-5 and CCR-3 for fusion^{7,8}. Full-length SDF-1 α (1–67) inhibited infection by the 89.6 virus of canine thymocytes expressing LESTR, but there was little effect on infection of CCR-5-transfected cells (Fig. 3c). These findings indicate that the HIV-1 inhibitory effect of SDF-1 is specific for LESTR. Furthermore, the experiments demonstrate that SDF-1 inhibits not only laboratory-adapted strains but also primary HIV-1 isolates that use LESTR for entry into permissive cells.

To determine whether SDF-1 would inhibit infection of primary cells, peripheral blood mononuclear cells (PBMCs) were used as target cells in the *env*-complementation assay. Full-length SDF-1 α (1–67), but not truncated SDF-1 α (6–67), inhibited the early phase of PBMC infection by virus pseudotyped with the envelope of the HXBc2 HIV-1 strain (Fig. 4).

Our search for the receptor for SDF-1 surprisingly led to the same orphan receptor as had been identified to mediate fusion of T-tropic HIV-1 strains with CD4-bearing cells⁶. We not only provide evidence that SDF-1 is a biological ligand for LESTR, but also that it specifically inhibits LESTR-mediated infection by both laboratory-adapted as well as primary T-tropic HIV-1 strains. Moreover, the latter experiments support the evidence that SDF-1 binds LESTR. Now that this receptor has been found to bind a CXC-chemokine, it can be designated CXC-chemokine receptor-4 (CXCR-4). The IL-8 receptor A and B and a recently described receptor for the monokine induced by gamma interferon (Mig) and the interferon-inducible protein 10 (IP-10) (ref. 20) represent the other members of the CXC-chemokine receptor family.

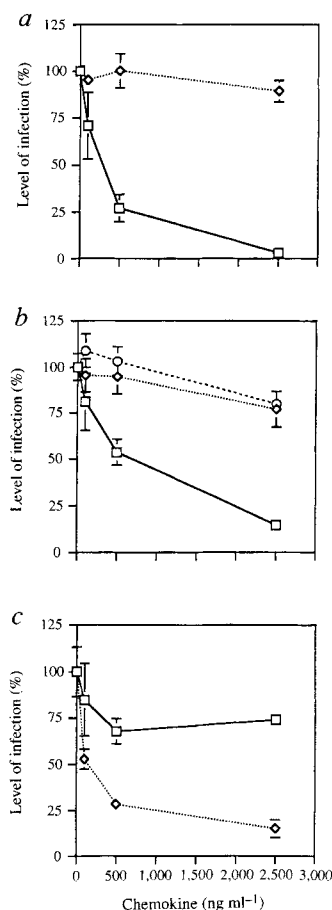
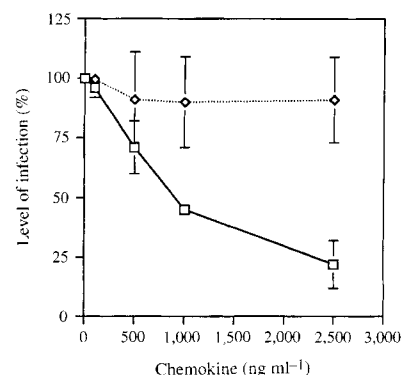


FIG. 3 Inhibition of LESTR-mediated infection of recombinant HIV-1 by full-length SDF-1 α (1–67). a, HeLa-CD4 cells expressing human CD4 were incubated with recombinant HIV-1 virus pseudotyped with the envelope glycoproteins of the laboratory-adapted HXBc2 isolate in the presence of different concentrations of full-length SDF-1 α (1–67) (squares) and N-terminally truncated SDF-1 α (6–67) (diamonds)^{7,19}. Results are indicated as percentage of chloramphenicol conversion observed in the absence of chemokine. Results with s.d. of three independent experiments are shown. b, Cf2Th-CD4 canine thymocytes transfected with either LESTR (squares, diamonds) or CCR-5 (circles) were incubated with recombinant HIV-1 virus pseudotyped with the envelope glycoproteins of either HXBc2- (diamonds, squares) or ADA- (circles) isolates. Full-length SDF-1 α (1–67) (squares, circles) and truncated SDF-1 α (6–67) (diamonds) were added at the indicated concentrations. Error bars show the range of duplicate experiments. c, Cf2Th-CD4 canine thymocytes transfected with either LESTR (diamonds) or CCR-5 (squares) were incubated in the presence of the indicated concentrations of SDF-1 α (1–67) with recombinant HIV-1 virus pseudotyped with the envelope glycoproteins of the dual-tropic primary HIV-1 isolate 89.6. Error bars show the range of duplicate experiments.

FIG. 4 Inhibition by SDF-1 α (1–67) of infection of PBMCs by recombinant HIV-1. PHA-stimulated PBMCs were incubated with CAT-expressing recombinant HIV-1 virus pseudotyped with the HXBc2 envelope in the presence of different concentrations of full-length SDF-1 α (1–67) (squares) and N-terminally truncated SDF-1 α (6–67) (diamonds). CAT activity was measured 60 h after infection. Results from duplicate wells of two independent experiments are indicated as percentage of chloramphenicol conversion observed in the absence of chemokine.



Although it is intriguing that SDF-1 is the most efficacious chemoattractant for T lymphocytes identified so far, and its receptor is fusogenic for T-tropic HIV-1 isolates, both the functional importance of SDF-1 and the cell distribution of CXCR-4 are much broader. SDF-1 is chemoattractive for B lymphocytes (C.C.B. and T.A.S., unpublished data), is a pre-B-cell stimulatory factor⁴, and is chemoattractive for haematopoietic stem cells (A. Aiuti, I. J. Webb, C.C.B., T.A.S. and J. C. Gutierrez-Ramos, in preparation). In agreement with this, mice genetically deficient in SDF-1 lack B cells and have myelopoiesis in the fetal liver but not in the bone marrow²¹. SDF-1 mRNA is well expressed in many non-haematopoietic tissues including the heart, brain, lung, kidney and liver^{3,12}. Similarly, CXCR-4 mRNA is well expressed in non-haematopoietic tissues including the brain, heart, kidney and lung^{15,22}; indeed, CXCR-4 is by far the most widely expressed of the functional chemokine receptors in non-haematopoietic cells. An important function for SDF-1 in non-haematopoietic cells is illustrated by perinatal lethality and failure of the ventricular septum of the heart to close in animals genetically deficient in SDF-1 (ref. 21). The high conservation of SDF-1 is mirrored in its receptor, which is 93% identical between human and bovine CXCR-4 (refs 8, 22). Because of the multiple biological functions, it will be important to determine whether SDF-1 binds to receptors other than CXCR-4, or whether there are additional ligands for CXCR-4.

Both the fusogenic receptors and their biological ligands are now known for most strains of HIV-1. M-tropic HIV-1 isolates use the chemokine receptors CCR-5 and CCR-3 as cofactors for membrane fusion⁷⁻¹¹. Chemokines that bind CCR-5 and CCR-3 have been shown to inhibit entry of primary M-tropic but not T-tropic HIV strains^{7,10,23}. Similarly, we have shown here that entry of a T-tropic strain that uses CXCR-4 is inhibited by the ligand SDF-1. Some primary HIV-1 isolates are T-tropic and use CXCR-4 (refs 7, 8), but most are M-tropic and use CCR-5 (refs 7-11); some primary isolates also use CCR-3 (refs 7, 8). Monocytes respond chemotactically to SDF-1 just as well as do T lymphocytes⁵, and express mRNA for CXCR-4 (ref. 18), so factors in the virus-host relationship other than ability to infect monocytes may determine the use of CCR-5 by most primary isolates. T-tropic HIV-1 strains can be found in HIV-1-infected individuals late in the course of infection. They induce the formation of syncytia in CD4⁺ cell lines, infect PBMCs faster, and replicate in these cells at a higher rate than M-tropic primary isolates^{24,25}. Occurrence of T-tropic primary HIV isolates in infected individuals is associated with CD4⁺ cell decline and progression to AIDS, and may suggest a detrimental role of these isolates in AIDS pathogenesis. The finding that SDF-1 attracts far more T lymphocytes than other chemokines⁵ suggests that CXCR-4 may be an abundant receptor on T cells, and may explain the use of this fusogenic receptor by T-tropic HIV-1. Why receptors for chemokines have been subverted by HIV-1 to trigger cell fusion and the molecular mechanisms that trigger fusion are intriguing subjects. The chemokines that bind these receptors will be important tools for HIV research; whether the chemokines, or antagonists of chemokine receptors, will prove useful in the fight against HIV is an important topic for further research. □

Methods

Chemokines and chemotaxis assay. Murine SDF-1 α was purified from supernatants of the murine bone-marrow stromal cell line MS-5 (ref. 5). Synthetic human SDF-1 α (1-67) and SDF-1 α (6-67) were generated using tBoc chemistry on an Applied Biosystems 430A peptide synthesizer²⁶. Proteins were folded by air oxidation and purified using at least two steps of reverse-phase high-performance liquid chromatography (RP-HPLC)²⁶. The molecular mass for each protein was determined by electrospray mass spectrometry on a API-3 triple quadrupole mass spectrometer (SCIEX, Thornhill, Ontario). The average masses $\pm$ s.d., with the calculated mass minus the measured mass in parentheses, were SDF-1 α (1-67) 7,831.63 $\pm$ 0.76 (-0.32) and SDF-1 α (6-67) 7,306.39 $\pm$ 0.45 (0.25). Lymphocyte chemotaxis assays were performed as described⁵. Human peripheral blood lymphocytes were obtained from healthy donors by Ficoll-Histopaque separation. Monocytes were removed by two 30-

min steps of plastic adherence. Cells (5×10^5) in 100 μ l RPMI-1640 medium, 0.25% human serum albumin (HSA), were added to the top chamber of a 5- μ m pore polycarbonate Transwell culture insert (Costar, Cambridge, MA) and incubated with the indicated concentrations of protein for 3 h. Transmigrated cells were counted with a FACScan (Becton Dickinson, San Jose, CA) using scattergates for lymphocytes for 20 s at 60 μ l min⁻¹. For chemotaxis of LESTR-transfected CHO cells, a modified Boyden-chamber chemotaxis assay was used²⁷. LESTR-transfected CHO cells were resuspended in RPMI 1640 medium containing 0.25% HSA and added in quadruplicate (15×10^3 cells per well) on top of a PVP-free, 8- μ m pore polycarbonate membrane previously coated with 15 μ g ml⁻¹ fibronectin/PBS (Sigma) in a microchemotaxis chamber (Neuro Probe, Cabin John, MD). Transmigrated cells were stained and cells visible in a $\times 20$ high-power field were counted.

Calcium fluorimetry. Untransfected CHO cells and the stable LESTR-transfected CHO cell line 1C2 were grown as described¹⁸. For Ca²⁺-mobilization studies, 10^7 cells were incubated in 1 ml loading buffer containing 136 mM NaCl, 48 mM KCl, 1 mM CaCl₂, 1 mg ml⁻¹ glucose and 20 mM HEPES, pH 7.4, with 5 nmol Fura-2 AM (Molecular Probes, Eugene, OR) for 20 min at 37 °C. The loaded cells were centrifuged, resuspended in fresh loading buffer, added to a stirred cuvette (2×10^6 in 500 μ l) and inserted into a Hitachi F2000 spectrometer. Chemokines were added in a volume of 5 μ l to a final concentration of 1 μ g ml⁻¹ at the indicated time points. Intracellular calcium concentrations were calculated using the Hitachi F2000 Intracellular Cation Measurement Program.

Envelope-complementation assay. Envelope-complementation assays were performed as described earlier^{7,19}. Recombinant virus was produced in HeLa cells by cotransfection with 15 μ g pHXBH10 Δ envCAT and 3 μ g pSVIIenv. HeLa-CD4 cells (clone 1022) were obtained from the NIH AIDS Research and Reference Reagent Program. Cf2Th canine thymocytes were transfected in duplicate by the calcium phosphate method with 10 μ g of pCD4 plasmid and 25 μ g of pCDNA3 expressing either LESTR or CCR-5, and were used approximately 60 h after transfection⁷. Approximately 8 h after replating, cells were incubated in 1 ml medium for 90 min at 37 °C with different concentrations of the indicated chemokine. Medium was removed and the cells were infected by incubation with recombinant virus (5,000 reverse transcriptase units for HXBc2-HeLa experiments and 20,000 reverse transcriptase units for HXBc2/ADA/89.6-Cf2Th experiments) pseudotyped with the respective envelope proteins in 1 ml of medium containing the original chemokine concentration. After 12 h at 37 °C, the medium was removed and fresh medium without chemokine was added. After an additional 48 h at 37 °C, the cells were lysed and CAT activity was measured. For assays with PBMCs, recombinant virus was produced in COS-1 cells by cotransfection with 10 μ g CMV Δ p1 Δ envpApu/vp, 5 μ g CMVenv Δ Xho and 7 μ g v653 RtatpC (HXBc2) as described^{28,29}. PBMCs were obtained from healthy donors by Ficoll-Hypaque separation and grown for 72 h with 1 μ g ml⁻¹ PHA (Murex Diagnostics, Dartford), rhuIL-2 (Becton Dickinson) was added to a final concentration of 20 U ml⁻¹ 18 h before infection. PBMCs (5×10^6 in experiment 1, and 1.5×10^6 in experiment 2) were incubated in 0.5 ml medium for 90 min at 37 °C with different concentrations of the indicated chemokine. Subsequently, recombinant virus (35,000 and 20,000 reverse transcriptase units for experiments 1 and 2, respectively), as well as the original chemokine concentration, were added in 1 ml. After 12 h at 37 °C the cells were washed and fresh medium without chemokine was added. After an additional 48 h at 37 °C the cells were lysed and used for measurement of CAT activity. Results are expressed as the percentage of chloramphenicol conversion observed in the absence of chemokine. The average chloramphenicol conversion for the experiments with the HXBc2 virus was 36% in HeLa cells, 20% in Cf3Th cells, and 22% in PBMCs.

Received 14 July; accepted 9 August 1996.

1. Baggiolini, M., Dewald, B. & Moser, B. *Adv. Immunol.* **55**, 97-179 (1994).
2. Springer, T. A. *Annu. Rev. Physiol.* **57**, 827-872 (1995).
3. Tashiro, K. et al. *Science* **261**, 600-603 (1993).
4. Nagasawa, T., Kikutani, H. & Kishimoto, T. *Proc. Natl. Acad. Sci. USA* **91**, 2305-2309 (1994).
5. Bleul, C. C., Fuhrbrigge, R. C., Casasnovas, J. M., Aiuti, A. & Springer, T. A. *J. Exp. Med.* (in the press).
6. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872-877 (1996).
7. Choe, H. et al. *Cell* **85**, 1135-1148 (1996).
8. Doranz, B. J. et al. *Cell* **85**, 1149-1158 (1996).
9. Dragic, T. et al. *Nature* **381**, 667-673 (1996).
10. Deng, H. et al. *Nature* **381**, 661-666 (1996).
11. Alkhatib, G. et al. *Science* **272**, 1955-1958 (1996).
12. Shirozu, M. et al. *Genomics* **28**, 495-500 (1995).
13. Roth, S. J., Carr, M. W., Rose, S. S. & Springer, T. A. *J. Immunol. Methods* **100**, 97-116 (1995).
14. Herzog, H. et al. *Proc. Natl. Acad. Sci. USA* **89**, 5794-5798 (1992).
15. Federspiel, B. et al. *Genomics* **16**, 707-712 (1993).
16. Jazin, E. E. et al. *Regul. Pept.* **47**, 247-258 (1993).
17. Nomura, H., Nielsen, B. W. & Matsushima, K. *Int. Immunol.* **5**, 1239-1249 (1993).
18. Loetscher, M. et al. *J. Biol. Chem.* **269**, 232-237 (1994).
19. Helseth, E. et al. *J. Virol.* **64**, 2416-2420 (1990).
20. Loetscher, M. et al. *J. Exp. Med.* (in the press).
21. Nagasawa, T. et al. *Nature* **382**, 635-638.
22. Rimland, J. et al. *Mol. Pharmacol.* **40**, 869-875 (1991).
23. Cocchi, F. et al. *Science* **270**, 1811-1815 (1995).
24. Asjo, B. et al. *Lancet* (ii), 660-662 (1986).
25. Schuitemaker, H. et al. *J. Virol.* **65**, 356-363 (1991).
26. Clark-Lewis, I. et al. *Biochem. J.* **30**, 3128-3135 (1991).

27. Yenush, L., Kundra, V., White, M. F. & Zetter, B. R. *J. Biol. Chem.* **269**, 100–104 (1994).
 28. Parolin, C., Dorfman, T., Palú, G., Gottlinger, H. & Sodroski, J. *J. Virol.* **68**, 3888–3895 (1994).
 29. Parolin, C., Taddeo, B., Palú, G. & Sodroski, J. *J. Virol.* **222**, (in the press).

CORRESPONDENCE and requests for materials should be addressed to T.A.S. (e-mail: collins@chr.med.harvard.edu).

ACKNOWLEDGEMENTS. We thank B. Moser for the untransfected CHO cells and the stable LESTR-transfected CHO cell line 1C2; B. Chesebro for the HeLa-CD4 (clone 1022) line; and the NIH AIDS Research and Reference Reagent Program for providing these cells. This work was supported by grants from the NIH (T.A.S. and J.S.) and fellowships from the Deutsche Forschungsgemeinschaft (C.C.B.) and the Institute Superiore di Sanita, Rome, Italy (C.P.), and gifts from the late William McCarty-Cooper, G. Harold and Leila Y. Mathers Charitable Foundation, and from the Friends 10.

The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1

Estelle Oberlin*, Ali Amara*, Françoise Bachelier*, Christine Bessia*, Jean-Louis Virelizier*, Fernando Arenzana-Seisdedos*, Olivier Schwartz†, Jean-Michel Heard†, Ian Clark-Lewis‡, Daniel F. Legler§, Marcel Loetscher§, Marco Baggiolini§ & Bernhard Moser§

*Unité d'Immunologie Virale and †Laboratoire Rétrovirus et Transfert Génétique, Institut Pasteur, 75724 Paris Cedex 15, France
 ‡Biomedical Research Centre, University of British Columbia, 2222 Health Sciences Mall, Vancouver, British Columbia, V6T1Z3, Canada
 §Theodor-Kocher Institute, University of Bern, PO Box 3000, Bern 9, Switzerland

A PUTATIVE chemokine receptor that we previously cloned and termed LESTR¹ has recently been shown to function as a co-receptor (termed fusin) for lymphocyte-tropic HIV-1 strains². Cells expressing CD4 became permissive to infection with T-cell-line-adapted HIV-1 strains of the syncytium-inducing phenotype after transfection with LESTR/fusin complementary DNA. We report here the identification of a human chemokine of the CXC type, stromal cell-derived factor 1 (SDF-1), as the natural ligand for LESTR/fusin, and we propose the term CXCR-4 for this receptor, in keeping with the new chemokine-receptor nomenclature. SDF-1 activates Chinese hamster ovary (CHO) cells transfected with CXCR-4 cDNA as well as blood leukocytes and lymphocytes. In cell lines expressing CXCR-4 and CD4, and in blood lymphocytes, SDF-1 is a powerful inhibitor of infection by lymphocyte-tropic HIV-1 strains, whereas the CC chemokines RANTES, MIP-1 α and MIP-1 β , which were shown previously to prevent infection with primary, monocyte-tropic viruses³, are

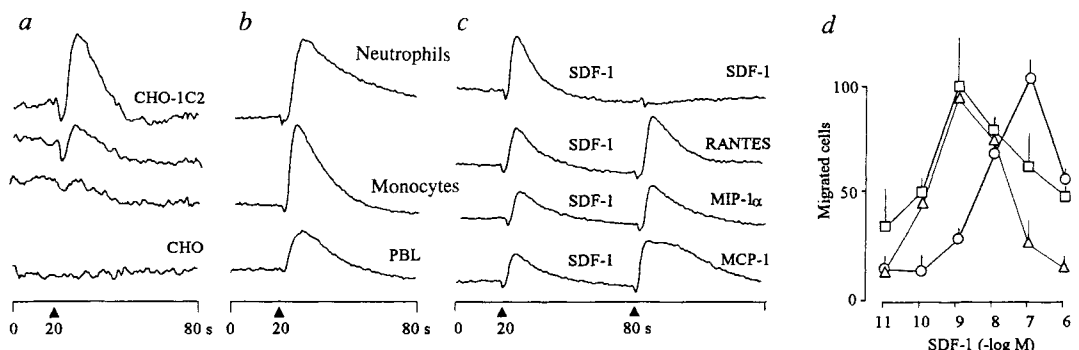
inactive. In combination with CC chemokines, which block the infection with monocyte/macrophage-tropic viruses, SDF-1 could help to decrease virus load and prevent the emergence of the syncytium-inducing viruses which are characteristic of the late stages of AIDS⁴.

LESTR (leukocyte-expressed seven-transmembrane-domain receptor) is an orphan receptor with structural similarity to chemokine receptors. Despite extensive testing of a large number of chemokines, the ligand for LESTR remained elusive¹. Murine SDF-1 was described as a factor that is produced by bone-marrow stromal cells and shown to induce proliferation of B-cell progenitors^{5,6} as well as recruitment of T cells⁷. The human homologue, which was cloned subsequently, is virtually identical to murine SDF-1 (see Methods). SDF-1 is a CXC chemokine with the typical four-cysteine motif and the first two cysteines separated by one amino acid⁸.

When human SDF-1 was tested on the CHO-1C2 clone which stably expresses LESTR, a transient rise of cytosolic free Ca²⁺ ([Ca²⁺]_i) was observed (Fig. 1a). This response, which is characteristic of the action of chemokines on blood leukocytes, was not observed with parental CHO cells. Other chemokines, including RANTES (for regulation-upon-activation, normal T expressed and secreted) macrophage inflammatory protein (MIP), MIP-1 α and MIP-1 β , were not active. Monocytes, neutrophils and phytohaemagglutinin (PHA)-activated peripheral-blood lymphocytes (PBLs) were also stimulated by SDF-1, as shown by [Ca²⁺]_i changes and chemotaxis (Fig. 1b, d). Real-time recordings of Ca²⁺ mobilization after sequential stimulation are a reliable way to assess receptor usage by chemokines⁸. Stimulation with a chemokine (at saturating concentrations) causes receptor desensitization, and no response is observed when the cells are restimulated within a short time by a chemokine acting on the same receptor. As shown in Fig. 1c, monocytes stimulated with SDF-1 remained fully responsive to subsequent stimulation with MCP-1, RANTES or MIP-1 α , and when applied first, none of these chemokines affected the [Ca²⁺]_i changes induced by SDF-1 as the second stimulus (not shown), indicating that SDF-1 does not share receptors with the three CC chemokines. The functional responses and the desensitization results show that LESTR is selective for SDF-1, the only ligand identified so far. The structure of receptor and ligand being known, we propose to rename LESTR/fusin as CXCR-4, in keeping with the receptor nomenclature established at the 1996 Gordon Research Conference on chemotactic cytokines.

The effects of SDF-1 on HIV-1 infection were tested first on HeLa cells transfected with CD4 and carrying an integrated HIV-1 long terminal repeat (LTR)-driven reporter gene, *lacZ* from *Escherichia coli*, which is induced by the HIV-1-encoded transactivating protein Tat. The induction of β -galactosidase, the product of *lacZ*, reflects transactivation by Tat and therefore HIV-1 infection. Like leukocytes and several tissue cells, the human epithelial carcinoma cell line HeLa constitutively expresses CXCR-4 (refs 2, 9). The cells were infected with two

FIG 1. SDF-1 is the ligand for CXCR-4 and is active on blood monocytes, neutrophils and on PHA-activated PBLs. a, SDF-1 induced [Ca²⁺]_i changes in the CHO-1C2 clone expressing CXCR-4 (100, 10, 1 nM, from top to bottom) and in parental CHO cells (100 nM). b, SDF-1-mediated [Ca²⁺]_i changes in human neutrophils, monocytes and PHA-activated PBL. SDF-1, RANTES, MIP-1 α and MCP-1 were used at 100 nM. c, Receptor desensitization in monocytes as assessed by [Ca²⁺]_i changes. d, *In vitro* chemotaxis of blood monocytes



(squares), neutrophils (circles) and PHA-activated PBLs (triangles) in response to SDF-1.

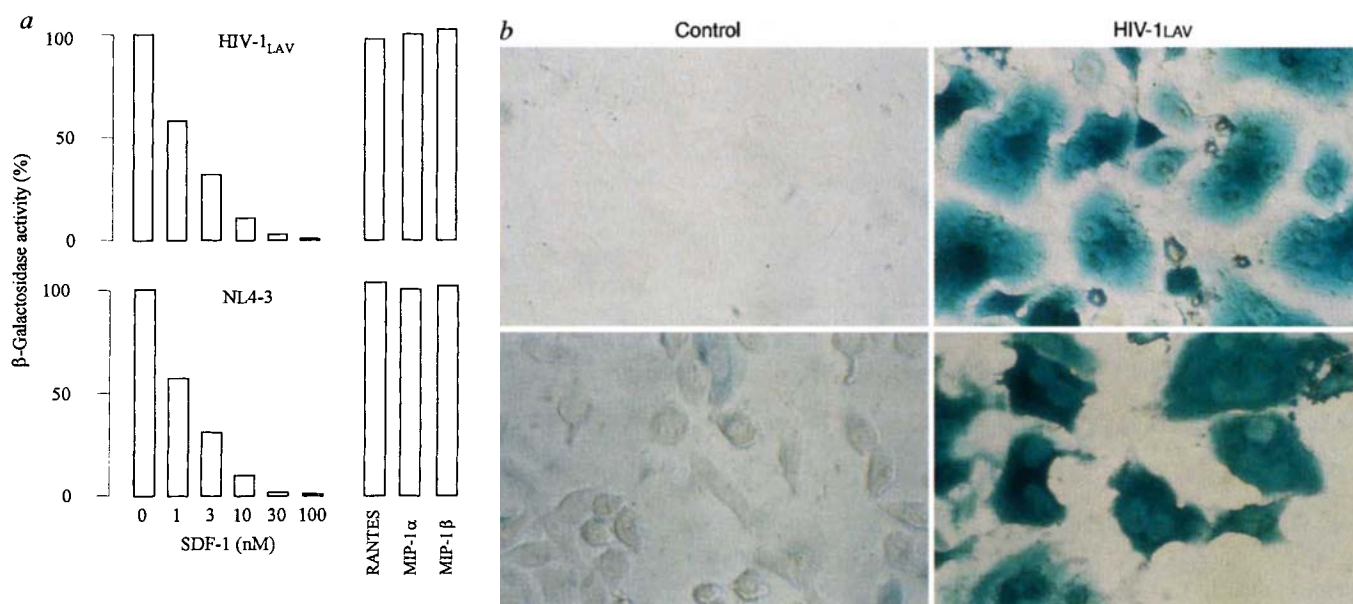


FIG. 2 SDF-1 inhibits infection of CD4⁺ HeLa cells by HIV-1 SI strains. **a**, SDF-1 (1–100 nM) but not RANTES, MIP-1 α or MIP-1 β (all 100 nM) inhibited infection by HIV-1_{LAV} and NL4-3. Mean per cent activity of triplicate values are shown; the data are representative of four similar experiments. **b**,

Formation of syncytia is inhibited by SDF-1 but not by RANTES. Cells infected with HIV-1_{LAV} and cultured for 24 h were fixed with 0.5% glutaraldehyde, and syncytia formation was visualized by staining for β -galactosidase.

viruses, HIV-1 Lai, LAV isolate (HIV-1_{LAV}) or the HIV-1 clone NL4-3, which harbours the LAV envelope (Env) glycoprotein¹⁰. SDF-1 was a very potent inhibitor of infection by both viruses, as shown by the decrease in β -galactosidase expression, which amounted to about 50, 70 and >95% at 1, 3 and 30 nM SDF-1, respectively (Fig. 2a). By contrast, no inhibition was observed when SDF-1 was replaced by 100 nM RANTES, MIP-1 α or MIP-1 β . These chemokines recognize the receptors CCR-5 and in part, CCR-1 and CCR-3 (refs 11–15), but they do not recognize CXCR-4 (Fig. 1). The dramatic inhibition of infection and syncytia formation by SDF-1 and the lack of activity by RANTES are shown in Fig. 2b. No expression of β -galactosidase was seen in infected cells cultured in the presence of the reverse-transcriptase inhibitor azidothymidine (AZT), indicating that the reporter gene was only induced upon infection (data now shown). To verify that the action of SDF-1 depends on the presence of CXCR-4, the receptor-expressing CHO-1C2 cells were transfected with CD4 and infected with NL4-3 carrying a luciferase reporter gene (kindly provided by V. Planellas). As in HeLa cells, SDF-1 inhibited the expression of the reporter gene, indicating inhibition of infection by NL4-3.

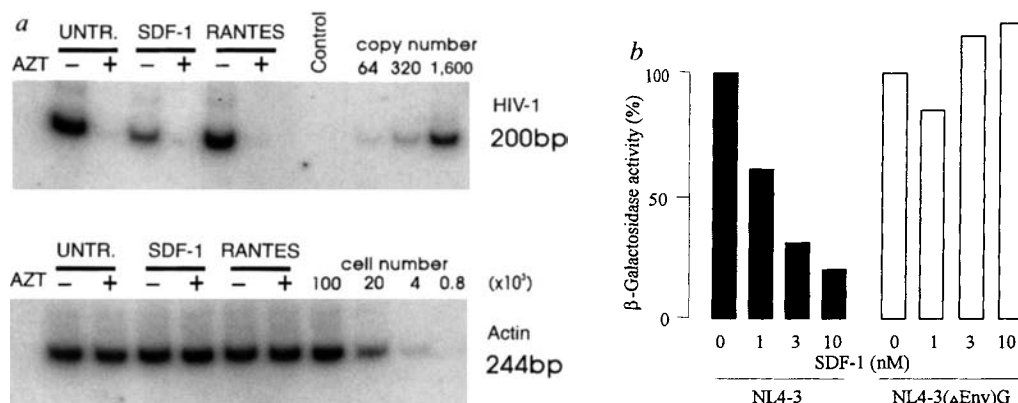
The inhibitory effect of SDF-1 was also studied by monitoring newly reverse-transcribed HIV-1 proviral DNA generated after

viral entry, using the polymerase chain reaction (PCR) technique. Extracts of CD4⁺ HeLa cells collected 6 hours after addition of HIV-1_{LAV} contained specific proviral DNA. In the presence of SDF-1, the appearance of proviral DNA was drastically reduced, whereas no effect was observed in the presence of RANTES. No transcripts were obtained when the reverse transcriptase inhibitor AZT was added to the cultures (Fig. 3a).

It was important to assess whether SDF-1 inhibits infection by competing with the virus for binding to CXCR-4, or by interfering with a post-entry event. For this purpose, we used an HIV-1 NL4-3 pseudotype expressing the G glycoprotein of vesicular stomatitis virus (VSV) instead of the LAV Env protein¹⁶ (kindly provided by A. Miyahara). NL4-3 (Δ Env)G readily infects cells through the interaction of the G glycoprotein with membrane phospholipids, and does not depend on HIV-specific receptors. As shown by the expression of β -galactosidase, SDF-1 had no effect on the replication of NL4-3 (Δ Env)G, in contrast to the marked concentration-dependent inhibition obtained for NL4-3 (Fig. 3b). These results rule out an effect of SDF-1 on the expression of HIV-1 proviral DNA, and suggest that SDF-1 inhibits HIV infection at stages that precede reverse transcription and presumably interferes with the interaction of the virus with CD4 and CXCR-4.

Upon activation, human peripheral blood mononuclear cells

FIG. 3 SDF-1 blocks entry of HIV-1_{LAV} in CD4⁺ HeLa(lacZ) cells. **a**, PCR amplification of retrotranscribed proviral DNA. The effect of SDF-1 or RANTES (both 100 nM) or buffer alone (UNTR.) on the generation of NL4-3 proviral DNA in CD4⁺ HeLa(lacZ) was examined. AZT (50 μ M) was added (+) or excluded (–) during infection to verify the retroviral origin of the PCR template. PCR products were analysed by agarose-gel electrophoresis and autoradiography. Three experiments were performed with similar outcome. **b**, Lack of inhibition of post-entry events by SDF-1 after infection by NL4-3 (Δ Env)G. Data are representative of four different experiments.



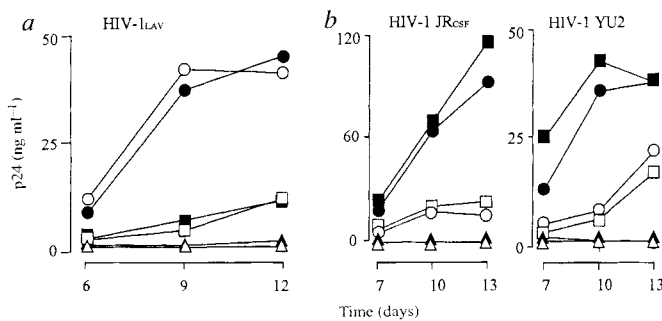


FIG. 4 SDF-1 is selective for lymphocyte-tropic HIV-1. **a**, Inhibition of HIV-1_{LAV} infection in PBMCs by SDF-1. 100 nM SDF-1 was added during and after infection (△, ▲), after infection (□, ■), or omitted (○, ●). Cultures with PBMCs from two different donors (open and filled symbols) were continued for 6 to 12 days, and p24 production assayed. **b**, No inhibition of HIV-1 JR_{SF} of HIV-1 YU2 infection in PBMCs by SDF-1. PBMCs from two different donors (open and filled symbols) were cultured for 7 to 13 d in the presence of 100 nM SDF-1 (○, ●), 100 nM RANTES (△, ▲) or in the absence of chemokine (□, ■), and p24 production assayed.

(PBMCs) support replication of HIV-1 *in vitro*. To test the inhibitory activity of SDF-1, PBMCs were activated by PHA before infection with HIV-1_{LAV} and the accumulation of soluble HIV-1 p24 protein was measured. When SDF-1 was present from the time of infection until the end of the experiment (day 12), protection was virtually complete (>98%), and a high degree of protection was observed even when SDF-1 was added after the period of infection (Fig. 4a). Infection with monocyte/macrophage-tropic HIV-1 (JR_{SF} or YU2), by contrast, was inhibited by RANTES but was not affected by SDF-1 (Fig. 4b).

The clear-cut discrimination of the mechanisms of entry by viruses with different tropism is highlighted by the inhibition of lymphocyte-tropic infection by SDF-1 and monocyte/macrophage-tropic infection by RANTES, MIP-1 α and MIP-1 β . This may be important when considering the progression of AIDS. Through its remarkable variability, the HIV-1 genome is likely to adapt the Env protein to different chemokine receptors. After years of asymptomatic infection with monocyto-tropic quasispecies viruses, strains may emerge that accelerate disease progression by rapid replication in leukocytes expressing CD4 with CXCR-4 as co-receptor. HIV adaptation to entry mediated by CXCR-4 into cells must represent a major biological advantage for the virus, which would find a much broader repertoire of infectable cells. In fact, CXCR-4 is constitutively expressed at high levels in PBMC¹, as well as in some tissue cells^{19,17} where expression of CC chemokine receptors is low. It is therefore of major importance to find means to block infection by CXCR-4-adapted syncytium-inducing viruses. SDF-1 administration or induction may inhibit dissemination of such viruses and possibly prevent their emergence in patients with chronic HIV infection. □

Methods

Responses to SDF-1 in CXCR4 transfected CHO cells and blood leukocytes. The 67-amino-acid form of human SDF-1 (accession number U16752), which differs from murine SDF-1 α (refs 5–7) by a single substitution (Val instead of Ile at position 18) and deletion of the C-terminal Lys, was prepared by chemical synthesis¹⁸. Neutrophils¹⁹, monocytes²⁰ and PBLs²¹ were prepared according to established methods. PBLs were stimulated for 72 h with 1 μ g ml⁻¹ PHA. Real-time recordings of [Ca²⁺]_i changes in CXCR-4-transfected CHO cells¹ and blood leukocytes loaded with Fura2/AM were performed as described²². Receptor desensitization was tested in monocytes by monitoring [Ca²⁺]_i changes upon repeated chemokine stimulation (100 nM) at 60-s intervals as described²⁰. Chemotaxis was performed in 48-well chambers with 10⁵ cells per well using bare (neutrophils and monocytes) or collagen-coated (PBLs) polyvinylpyrrolidone-free polycarbonate membranes with 5- μ m pores²⁰. Migrated cells were counted in five randomly selected fields at 1,000-fold magnification after migration for 1 h (neutrophils and monocytes) or 3 h (PBLs).

Generation of HIV-1 particles. HIV-1_{LAV} and NL4-3 infectious supernatants

were obtained from infected CEMX174 and MT4 lymphoblastoid cell lines, respectively. NL4-3 particles¹⁰ were generated by transient transfection of COS7 cells with cloned pNL4-3 DNA. NL4-3(Δ Env)G, a pseudotyped NL4-3 variant with a replacement of the Env protein with the G glycoprotein of VSV, was produced by co-transfection of COS7 cells with a plasmid encoding NL4-3 with a BglII–BglII detection in the Env gene (provided by V. Planellas) plus pCMV vector carrying the DNA for VSV G glycoprotein. HIV-1 JR_{SF}²³ and YU2²⁴ particles were obtained by transfection of COS7 cells with plasmids carrying full-length proviral DNA.

HIV-1 entry and cell-fusion assays. A CD4⁺ HeLa(lacZ) cell clone was generated in HeLa cells carrying a stably integrated lacZ gene under HIV-1 LTR control by stable transfection with a retroviral vector containing CD4 cDNA²⁵. 1.5 $\times$ 10⁴ CD4⁺ HeLa(lacZ) cells per well in 96-well plates were cultured for 18 h before infection with 250 μ l infectious supernatants (HIV-1_{LAV}, NL4-3 or NL4-3(Δ Env)G at 500, 50 and 1 ng ml⁻¹ HIV-1 p24, respectively) in the presence or absence of chemokines. After 24 h, β -galactosidase activity was determined in cell lysates from triplicate wells. After removing supernatants, cells were lysed in 100 μ l 60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 10 mM MgSO₄, 2.5 mM EDTA, 50 mM β -mercaptoethanol and 0.125% NP-40, and mixed with 100 μ l 80 mM sodium phosphate, pH 7.4, 10 mM MgCl₂, 50 mM β -mercaptoethanol and 6 mM chlorophenol red- β -galactopyranoside monosodium salt and incubated for 20 min at 37 °C before measuring absorbance at 540 nm.

PCR amplification of proviral DNA. 2 $\times$ 10⁶ cells in 4 ml were infected for 2 h with 2 ml NL4-3 infectious supernatant (100 ng ml⁻¹ HIV-1 p24). The cultures were washed 3 times, supplemented with 100 nM SDF-1, 100 nM RANTES or the vehicle, and cultured for an additional 6 h in the presence (+) or absence (–) of 50 μ M AZT. After extraction of total DNA²⁶, PCR was performed with primers for proviral DNA (3'-region primer, CTGTCGTCGAGAGCTCCTCTGG-3'; and 5'-region primer, GGCTAACTAGGGAACCCACTG-3'; end-labelled with ³²P) in 50 μ l 0.25 mM dNTPs, 50 mM NaCl, 25 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 100 mg ml⁻¹ BSA and 1.5 units of Taq DNA polymerase (Amersham) for 25 cycles. The primers correspond to nucleotide positions 496–516 (first nucleotide is at position 42 downstream of the transcription initiation site) and 695–672 (complementary sequence in gag) in the HIV-1 JR_{SF}. Primers to β -actin DNA (5'-region primer, GTGGGGCGCCCCAGGCACCA-3', and 3'-region primer, CGGTGGCCTTGGGGTTCAGGGGGG-3') yielding a 244-bp amplified product were used for standardization.

Infection of PBMCs. PBMCs from healthy blood donors were stimulated for 48 h with 1 μ g ml⁻¹ PHA (Wellcome) in RPMI 1640, 10% fetal bovine serum (FCS), and then infected for 2 h with 10–20 ng of HIV-1 p24 (10⁷ cells per infection in 4 ml) in the presence or absence of chemokines. Infected PBMCs were washed four times with PBS and cultured in RPMI 1640, 10% FCS, plus 20 ng ml⁻¹ IL-2 with or without chemokines. Cultures were refed every 3 d with IL-2 and chemokines, and soluble HIV-1 p24 was determined by ELISA (DuPont de Numours, France).

Received 17 July; accepted 12 August 1996.

- Loetscher, M., Geiser, T., O'Reilly, T., Zwaen, R., Baggiolini, M. & Moser, B. *J. Biol. Chem.* **269**, 232–237 (1994).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872–877 (1996).
- Cocchi, F. et al. *Science* **270**, 1811–1815 (1995).
- Weiss, R. A. *Science* **272**, 1885–1886 (1996).
- Tashiro, K. et al. *Science* **261**, 600–603 (1993).
- Nagasawa, T., Kikutani, H. & Kishimoto, T. *Proc. Natl Acad. Sci. USA* **91**, 2305–2309 (1994).
- Bleul, C. C., Fuhlbrigge, R. C., Casasnovas, J. M., Aiuti, A. & Springer, T. A. *J. Exp. Med.* (in the press).
- Baggiolini, M., Dewald, B. & Moser, B. *Adv. Immunol.* **55**, 97–179 (1994).
- Nomura, H., Nielsen, B. W. & Matsushima, K. *Int. Immunol.* **5**, 1239–1249 (1993).
- Adachi, A. et al. *J. Virol.* **59**, 284–291 (1986).
- Deng, H. K. et al. *Nature* **381**, 661–666 (1996).
- Dragic, T. et al. *Nature* **381**, 667–673 (1996).
- Alkhatib, G. et al. *Science* **272**, 1955–1958 (1996).
- Choe, H. et al. *Cell* **85**, 1135–1148 (1996).
- Doranz, B. J. et al. *Cell* **85**, 1149–1158 (1996).
- Yee, J. K. et al. *Proc. Natl Acad. Sci. USA* **91**, 9564–9568 (1994).
- Jazin, E. E. et al. *Regul. Pept.* **47**, 247–258 (1993).
- Clark-Lewis, I. et al. *Biochemistry* **30**, 3128–3135 (1991).
- Peveri, P., Walz, A., Dewald, B. & Baggiolini, M. *J. Exp. Med.* **167**, 1547–1559 (1988).
- Ugucioni, M., D'Apuzzo, M., Loetscher, M., Dewald, B. & Baggiolini, M. *Eur. J. Immunol.* **25**, 64–68 (1995).
- Colotta, F., Peri, G., Villa, A. & Mantovani, A. *J. Immunol.* **132**, 936–944 (1984).
- von Tschamer, V., Prod'homme, B., Baggiolini, M. & Reuter, H. *Nature* **324**, 369–372 (1986).
- Cann, A. J. et al. *J. Virol.* **64**, 4735–4742 (1992).
- Li, Y. et al. *J. Virol.* **66**, 6587– (1992).
- Clavel, F. & Chameau, P. *J. Virol.* **68**, 1179–1185 (1994).
- Jacque, J. M. et al. *J. Virol.* **70**, 2930–2938 (1996).

ACKNOWLEDGEMENTS. E.O. and A.A. contributed equally to this work. This study was partially supported by the Agence Nationale pour la Recherche sur le Sida (ANRS), the EU Concerted Action (project ROCIO), the Swiss National Science Foundation, and the National Institute of Health. Donor blood buffy coats were provided by the Swiss Central Laboratory Blood Transfusion Service, SRK. B.M. holds a career development award from the Prof. Max Clœtta Foundation, and I.C.-L. is recipient of a Scientist award from the MRC of Canada. E.B. and A.A. are supported by a fellowship from ANRS and Sidaction (France).

CORRESPONDENCE and requests for materials should be addressed to F.A.-S. (e-mail: farezan@pasteur.fr) or to B.M. (e-mail: bmoser@tki.unibe.ch).

Guide to authors of contributions to *Nature*

Nature is a weekly international journal covering all the sciences. It is intended for an interdisciplinary readership, so all manuscripts should be written clearly and simply. Contributors should pay particular attention to the fact that English is not the first language of many readers. Space is limited and competition severe, so brevity is highly valued. The length limits below are specified in terms of the number of pages taken up in *Nature*. An uninterrupted page of text contains about 1,300 words, so a typical Letter to *Nature*, occupying 2½ pages, can contain about 1,500 words of text and four modest display items (figures and/or tables) with brief legends. There are no formal limits on the number of display items, but extra display material must be accompanied by commensurate reductions in the length of the text. Length limits can, however, be varied upwards or downwards at the editors' discretion.

Nature publishes the following types of contribution:

Review Articles survey recent developments in a topical area of scientific research. They should not generally occupy more than 6 pages of *Nature*.

Progress articles are relatively informal reviews of particularly topical and fast-moving fields. They do not exceed 4 pages of *Nature*.

Contributors wishing to submit a Review Article or Progress should first send a one-page synopsis to the Reviews Coordinator, from whom guidelines are available on request.

Articles are reports of original scientific research whose conclusions represent a substantial advance in understanding of an important problem and are of broad general interest. They do not generally exceed 5 pages of *Nature* and have no more than 50 references. They have a 'heading' of up to 80 words, often rewritten by *Nature's* editors, which advertises the content of the paper in general terms and should not contain numbers, abbreviations or measurements unless central to the message of the paper. The introduction and summary should be contained in the first three or four paragraphs of the main text. Layout of methods, figures and legends is summarized below.

Letters to *Nature* are short reports of original research focused on an outstanding finding whose implications are of interest to nonspecialists. They should not generally exceed 2½ pages of *Nature*. They have a bold first paragraph of not more than 180 words summarizing the background, rationale, conclusions and main implications of the work, aimed particularly at nonspecialist readers. Letters should have no more than 30 references. Layout of methods, figures and legends is summarized below.

Preparation of Articles and Letters

All manuscripts should be written in English, typed double-spaced on one side of the paper only. Manuscripts submitted in electronic form cannot be considered.

Manuscripts should be accompanied by a brief covering letter from the corresponding author, containing a full postal address, telephone and fax numbers, and e-mail address.

Five copies of manuscripts and original figures are needed, together with two copies of the covering letter. Five copies of manuscripts by any of the authors that have been submitted or are in press elsewhere should also be included, clearly marked as such; failure to do so may prejudice consideration of the manuscript.

Unless otherwise instructed by the editor handling a manuscript, when sending revised or resubmitted manuscripts five copies are required, each accompanied by a copy of the authors' response to referees' and editors' comments on the earlier version.

Titles should not exceed three lines of 31 characters each (including spaces), and should avoid numerical values, acronyms, abbreviations and punctuation wherever possible.

References should be numbered sequentially as they appear in the text, tables and figure legends. Only papers that have been published or submitted to a journal should be in the numbered list: manuscripts in preparation should be mentioned in the text with a list of authors. Reference lists should not contain textual material, grant details or acknowledgements.

Acknowledgements must be brief and appear after the reference list. *Nature* does not publish grant or contribution numbers.

Figures should not be larger than 22 by 28 cm unless unavoidable. They should not contain more than one panel unless the assembled parts are logically connected. Each copy should be marked with the corresponding author's name and, when known, the manuscript reference number. One photocopy of the original figures should be provided. All original figures are returned when *Nature* cannot offer to

publish a manuscript, but one copy of the manuscript and a photocopy of the figures are kept in confidential files for three months and then destroyed. Detailed instructions for figure preparation are available from the production department, including details of acceptable digital formats.

Colour figures are welcome but a contribution towards the cost of reproduction is requested. Inability to pay this charge will not prevent *Nature* publishing figures where colour is essential.

Tables. Symbols and abbreviations in a table should be defined immediately below the table, followed by essential descriptive material, all in double-spaced text. Tables should each be presented on a separate sheet of paper.

Methods. Brief details of materials and methods can be given in the text. Otherwise, they should be described at the end of the text in a section entitled "Methods", subdivided by bold headings referring to the methods used. This section should not exceed 800 words and ideally should be shorter. Additional material can be deposited on *Nature's* Web site. In contributions with methods sections, figure legends should not exceed 150 words each, and should be confined to a description of the figure and the symbols used.

Materials. As a condition of publication, authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting datasets must be made available at the time of publication from the authors directly, by posting on *Nature's* Web site, by deposition in the appropriate public database or by distribution on the Internet. In the case of X-ray crystallographic coordinates only, public access may be delayed for up to one year after publication.

Articles and Letters should both contain a statement at the end of the text: "Correspondence and requests for materials to xxxxx", with an e-mail address if desired, followed by the accession numbers or Web site address for supporting datasets. The designated author need not necessarily be the senior author, but *Nature* will expect the named author to handle matters arising from published contributions.

Manuscripts can be submitted to the Editor at *Nature*, Porter's South, Crinan St, London N1 9XW or at *Nature*, 968 National Press Building, Washington, DC 20045-1938. Proofs should be returned by express mail to London. Fax numbers and e-mail addresses for all editorial offices are on *Nature's* masthead.

Other contributions

Nature publishes informal contributed material as follows:

Correspondence contributions are short comments on topical issues, anecdotal material or reactions of readers to nontechnical material published in *Nature* (mainly News, Briefings, Opinion, Book Review or Commentary articles). All are unsolicited; submissions by e-mail to corres@nature.com are preferred (accents on names and places should be clearly indicated).

Commentary articles are informal, editorial comments on topical issues of public interest that have some bearing on scientific research. They are often controversial. Unsolicited manuscripts are considered, but enquiries can be made to the Commentary Editor by e-mail (nature@nature.com) or fax as a one-page synopsis or informal letter before formal submission.

News and Views articles inform nonspecialist readers about new scientific advances, either as reported in published papers or in the form of meeting reports. Most are commissioned, but proposals can be made to the News and Views Editor in advance of publication of the paper or of the meeting concerned. Authors are not allowed to discuss their own work or work from elsewhere in their institution. Detailed guidelines are available on request.

Scientific Correspondence is a predominantly peer-reviewed section of the journal specializing in very brief reports of general interest and discussion of topical issues. Priority is given to contributions of fewer than 500 words. Detailed guidelines are available on request. All contributions are unsolicited, and should be submitted by mail, not e-mail. Copies of contributions submitted by fax should be sent by mail with original figures and a cover letter noting that the manuscript has been faxed.

Book Reviews are all commissioned by the Book Review Editor. Unsolicited contributions are not considered.

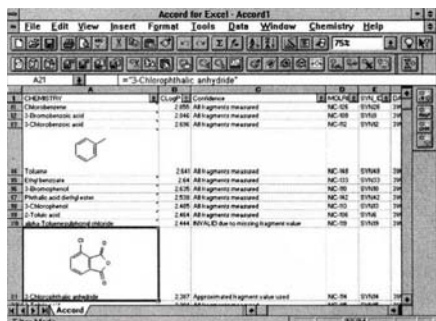
Nature's editors will request final copy for Commentary, News and Views, Scientific Correspondence and Book Reviews on disk or via e-mail, as appropriate.

A more detailed Guide to Authors is available on *Nature's* WWW sites at <http://www.nature.com> or <http://www.america.nature.com>

A proliferation of programs

Time once again to contact your software providers who are peddling such wares as a searchable chemical database, an Internet search engine, chemometrics software and virus protection software for networks.

Synopsys Scientific Systems and Daylight Chemical Information Systems have jointly announced the development of a new add-on for Synopsys' **chemical spreadsheet**, Accord for Microsoft Excel. This add-on provides users with the ability to calculate physicochemical properties, such as LogP and molar refractivity (MR), through a client/server linked to the Daylight tool server module. Accord provides scientists with a full range of chemical facilities to manage, analyse and search chemical 'objects' and associated



NAME	CONFIDENCE	LOGP	MR	STATUS
1-Chlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2-Dichlorobenzene	2.000	All agents measured	NC-01	ST-01
1,3-Dichlorobenzene	2.000	All agents measured	NC-01	ST-01
1,4-Dichlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3-Trichlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,4-Trichlorobenzene	2.000	All agents measured	NC-01	ST-01
1,3,5-Trichlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,4-Tetrachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,5-Tetrachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,6-Tetrachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,4,5-Tetrachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,3,4,5-Tetrachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,4,5-Pentachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,4,6-Pentachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,5,6-Pentachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,4,5,6-Pentachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,3,4,5,6-Pentachlorobenzene	2.000	All agents measured	NC-01	ST-01
1,2,3,4,5,6-Hexachlorobenzene	2.000	All agents measured	NC-01	ST-01

Add-on for Accord chemical spreadsheet.

data within the spreadsheet environment. It also provides an additional toolbar that allows a user to connect to the tool server and obtain CLogP or CMR values for molecules in a selected spreadsheet region. These options should allow a user to obtain supplementary information returned from the tool server, such as confidence values or whether a measured value is available. In addition, a chemical table can be filtered or sorted based on CLogP or CMR values.

Reader Enquiry No. 100

SciFinder version 2.0 from Chemical Abstracts Service (CAS) is designed to be an easy-to-use **research tool to assist in planning and optimizing chemical syntheses**. In addition to the capabilities available in the original release, the new version introduces features permitting scientists to search for reactions, find commercial source information for chemical substances, order documents from CAS, and more. According to the manufacturer, SciFinder can now execute a reaction substructure search, which reviews more than 1.2 million single-step reactions compiled from 1985 to the present. Chemical availability, prices and quantities for more than 370,000 products from a variety of North

American and European suppliers' chemical catalogues are now available. Other improvements permit customers to find regulatory compliance information, specify roles in a reaction search (for example, by indicating whether substances of interest are reagents or products), and communicate comments electronically to CAS. SciFinder is said to provide access to more than 15 million chemical substance records, nearly 14 million documents, information on commercially available chemical products, 2 million patents from 26 countries and 1.2 million chemical reactions. This information is updated daily or weekly.

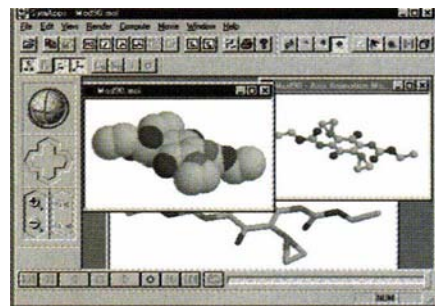
Reader Enquiry No. 101

Advanced ChemTech and Chemical Design have finalized an agreement to market a new version of Chemical Design's popular Chem-X **combinatorial chemistry software** designed specifically for use with Advanced ChemTech's multiple organic synthesis instrumentation. The new software package is designed to combine integrated informatics and high-throughput synthesis, providing capabilities for affordable lead generation and optimization. The new Windows-based software is said to simplify the synthesis of combinatorial libraries by automatically producing ready-to-run synthesis protocols for these multiple organic synthesizers. The software automatically generates all the reagent definitions and dispense-sequence protocols needed to execute a wide variety of combinatorial syntheses. The new software will be included with all purchases of Advanced ChemTech multiple organic synthesis instrumentation and offered as a software upgrade to current owners of Advanced ChemTech instruments. Customers can choose from multiple software configurations with varying library design capabilities. The Chem-X/Diverse system to be supplied with Advanced ChemTech's Model 496 MOS will feature lead generation and optimization capabilities, including diversity analysis and library design, R-group selection and biological data management.

Reader Enquiry No. 102

Structures and modelling

SoftShell now offers SymApps a new **three-dimensional, molecular-rendering program** designed for desktop visualiza-



Molecular rendering in three dimensions.

tion and publishing. Molecular drawings produced by the program are said to be ideal for handouts, presentations and publishing. The program features the ability to calculate symmetry point groups, to display the symmetry elements on-screen and can also calculate character tables for any symmetry point group. For electronic presentations and to aid in visualizing molecules, the program can create animated movies of symmetry operations. SymApps was designed as an OLE server application, which allows users to embed molecules inside OLE-compatible programs, such as Microsoft Word. It will open common three-dimensional file formats, including molfiles, Gaussian, Mopac, PDB and XYZ. Users can also copy and paste two-dimensional structures from ChemWindow DB, ChemWindow 3.0 or ISIS/Draw, and use the built-in MM calculations to convert structures to three dimensions. Version 1.0 requires Windows NT or Windows95, a fast 486 CPU (Pentium preferred), and 16 MB of RAM. SymApps retails for US\$499 and academic and student discounts are available.

Reader Enquiry No. 103

Tripos has announced the availability of ChemSpace, a **searchable chemical database** consisting of billions of chemicals. The database is based on proprietary technology for storing and searching vast numbers of chemical structures and their associated descriptive data. Searches proceed at effective rates above 100 million molecules per hour with access via World-Wide-Web-enabled interfaces. The database's speed and ease of use should increase the efficiency of compound discovery. Many pharmaceutical and biotechnology companies have, for example, implemented high-throughput screening

programs in their compound discovery procedures. ChemSpace offers the ability to search structurally all possible chemical compounds from a given set of known reactions and available reagents. In this way, it is said to provide a systematic and efficient approach to the design of libraries and compounds.

Reader Enquiry No. 104

Molecular Simulations (MSI) a provider of **molecular modelling software**, is making its technology available to a wider audience using the World Wide Web. WebLab, as it is called, is the company's new environment for molecular simulation and computational chemistry and biology. The manufacturer states that any desktop computer with a Web browser can be used to access MSI's collection of scientific software through the Internet or local intranets. Using this environment as a new vehicle to deploy existing and new software, MSI expects to reach more researchers and scientists who, until now, had no way to access the company's advanced modelling software. Users will gain access to MSI's technology, not just from workstations, but from Macintosh and Intel-based personal computers. WebLab Viewer, the first in a line of products, allows researchers to view chemical structures from an extensive library. The viewer, offered free of charge, can be downloaded from MSI's homepage beginning this week. (<http://www.msi.com/weblab>).

Reader Enquiry No. 105

CambridgeSoft has unveiled a new version of its **centralized Internet search engine** ChemFinder WebServer, providing its users with a means to locate chemical information anywhere on the Internet (<http://chemfinder.camsoft.com>). The engine searches chemical information by a number of criteria, including name, formula, molecular weight, CAS registry number and molecular structure. It then provides basic physical property information about the compounds, along with links to other Internet sites that have more information about each particular compound. An extensive synonyms database within ChemFinder

facilitates the use of popular names in addition to standardized chemical nomenclature. The engine also recognizes regional differences in chemical names and other common variances in naming systems. In addition to providing one of the largest listings of chemical information sites on the Internet, the program's master index is designed to transcend naming variations.

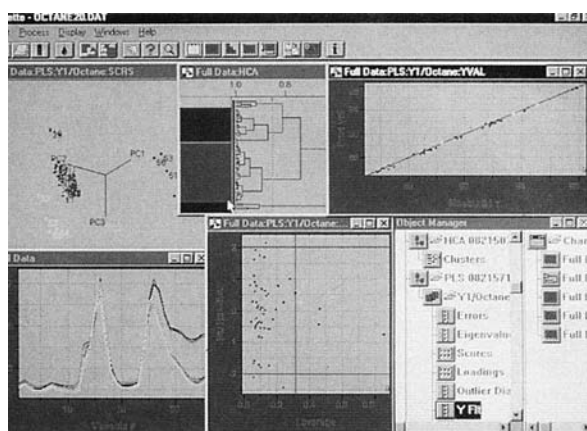
Reader Enquiry No. 106

Chemometrics

Galactic now offers a brochure covering the technical and practical aspects of its **PLSplus/IQ chemometrics package**. This new full-colour brochure describes how this software provides a complete solution for analysts looking to apply chemometric calibrations in methods, advanced diagnostics, pre-processing procedures and report plotting. This program is an add-on application for GRAMS/32 and GRAMS/386, which includes powerful multivariate analysis tools for building quantitative calibrations, as well as qualitative models for discriminant analysis. GRAMS/32 is an optimized, 32-bit software product that runs on 386, 486 or Pentium computers under Windows 3.1 and Windows95. A maths coprocessor and VGA display are required.

Reader Enquiry No. 107

Version 2.0 of Pirouette **chemometrics software** has been optimized for Windows NT and Windows95 and is now available from InfoMetrix. The program now supports drag-and-drop graphics, virtually unlimited data-set sizes and concurrent sessions. This new product is designed to bring a more comprehensive and dynamic project environment to the world of chemometrics. Management of data and results has been integrated into a tree-like structure to form an intuitive object management system. Moreover, an unlimited number of multivariate models can be



Pirouette chemometric software for Windows95 and NT.

stored and managed as part of the Pirouette data file. Batch analysis of an assortment of data subsets with several algorithms can be set up for unattended processing with full cross-validation included in the exploratory (PCA) algorithm. For tailored applications, a macro-driven set of commands allows custom sequences to run using the chemometrics processing engine.

Reader Enquiry No. 108

Statistics and graphics

Visual Numerics, a supplier of **decision support software tools for areas of data analysis and visualization**, has announced the availability of PV-Wave version 6.0 for both Microsoft Windows95 and Windows NT 3.51 operating systems. These versions should provide customers with the ability to develop and deliver the same application across a multi-vendor environment including UNIX and PC systems. Short development cycles, competitive pressures and the complexities associated in developing and delivering applications across multi-vendor environments are said to necessitate the development of new approaches to application development. The program offers complex data analysis and visualization, using a fourth generation programming language, which should reduce application development and deployment times. It now allows applications to be run under 32-bit Windows, while retaining full compatibility with the UNIX version.

Reader Enquiry No. 109

Jandel Scientific Software has introduced a new version of TableCurve 2D, giving more curve-fitting capabilities to this flexible **curve-fitting and equation discovery program**. This package can be used with Windows95, NT or 3.1 to locate best-fit equations automatically from a built-in set of over 3,600 functions. Users also have the option of entering their own functions for fitting and ranking. Engineers, scientists and economists

ADVERTISEMENT



Do you collect AIR MILES awards?

Do you lecture in the UK?

If so, please visit the

Chapman and Hall website at

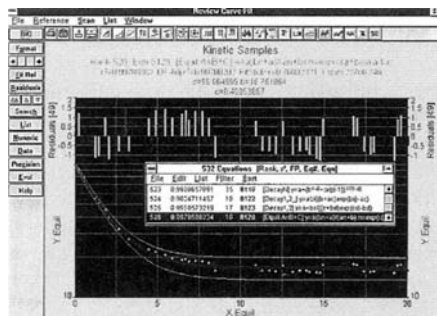
<http://www.chaphall.com>



AIR MILES and the Flying Boat logo are trademarks of AIR MILES International Holdings N.V.
AIR MILES awards are held and issued for use subject to AIR MILES Customer Terms and Conditions.

READER ENQUIRY NO. 76

should find the program useful for a wide range of applications, such as optimizing process control, generating calibration curves, creating blackbox models, modelling complex data, and fitting tabulated data. Version 4.0 features over 200 additional built-in equations and can now fit polynomials up to 20th order and rationale up to 10th/10th order, with matrix computations done to a full 38 significant digits. In total, 88 new polynomial equations and 110 new rational equations have



Jandel's TableCurve 2D software.

been added, as well as twenty-four new nonlinear peak functions and two transition models. Three non-parametric estimations for interpolation have been included: local regression estimation, spline fitting and smooth-spline estimation. The program can output more program codes than previous versions (Visual Basic, Fortran 90, and 80-bit C code), in addition to the language formats in previous versions.

Reader Enquiry No. 110

SPSS has released the Systat 6.0 **statistical software package** for Windows with improvements designed to make the system easier to use. A specification sheet is available that describes the capabilities of this system and the company's upgrade policy, which makes it possible to upgrade for a limited time not only from any earlier version of Systat but also from other statistical products. In addition, existing users who bought the Windows version after 1 December 1995 will receive their upgrades to version 6.0 at no charge. An order form accompanies the specification sheet, and can be used to order both upgrades and new copies.

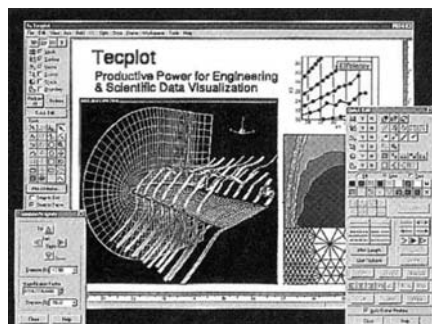
Reader Enquiry No. 111

MathSoft has announced the release of S-PLUS 3.4 for UNIX, an object-oriented, **data-analysis package** that is based on the 'S' language developed by AT&T Bell Labs (now Lucent Technologies). It is designed for use by statisticians and researchers, and includes new advanced statistical methodologies, expanded graphics display features, as well as an entirely new set of validation routines. Version 3.4 includes an abundance of data analysis methods, allowing statisticians to improve the analysis of

their data. Some new techniques include nonlinear, mixed-effects modelling and advanced cluster-analysis techniques, which are important tools for analysing multivariate and longitudinal data. Visualization methods are provided by new Trellis display methods and by the first release of hexagonal binning. Trellis includes new procedures for animation and colour draping. Moreover, this allows users to create multipanelled graphical displays of one or two variables, arranged according to the values of additional related variables. The displays in each panel can include histograms, scatter plots, dot plots, contour plots and wire-frame plots. Hexagonal binning is a tool for the visualization of large and dense bivariate plots. Users have flexible control over axes and aspect ratios.

Reader Enquiry No. 112

Amtec Engineering's new Tecplot 7.0 software should provide engineers and scientists with a broad set of **tools available for visualizing and plotting large amounts of data**. This latest version provides a new graphical user interface, animation and page layout. It should help users work productively with the data sets generated by numerical simulation (such as, computational fluid dynamics), statistical analysis, data acquisition, and other sources. The primary purpose of this program is to visualize and interact with complex data sets,



Tools for data plotting and visualization.

transforming thousands or millions of data points into a variety of understandable two- or three-dimensional visual images. A wide selection of viewing options includes wire-mesh plots, contour lines with flooded contours, vector fields, light-source shaded plots and numerous x/y plots. Tecplot's second purpose is to make it easy to annotate plots and adjust the amount of information within a plot to match the audience's familiarity with the data. For engineers and scientists there are multi-block, curvilinear grids that accurately model real-life objects, from aircraft wings to car engines to soil structures. It also offers unstructured finite-element grids, true three-dimensional surface and volumetric modelling, animation and data manipulation to customize results. The software lets

users interact with data in a point-and-click fashion and is designed to switch plot types and views quickly. Tecplot version 7.0 runs on most UNIX workstations (under Motif) and PCs running Windows (3.x/95/NT). In addition, Tecplot will soon be available on VMS, Linux, and Windows NT for DEC Alpha. Single-user pricing ranges from US\$995 to US\$3,195, depending on the platform and the type licence.

Reader Enquiry No. 113

Bibliographic software

Research Information Systems has taken full responsibility for managing Personal Bibliographic Software, a supplier of **bibliographic database software programs**. The product line includes ProCite, a PC-based personal bibliographic database management program; BiblioLinks, a set of programs used to import reference data from on-line-, CD-ROM-, and diskette-based services into ProCite; and the recently added Internet Enabler, which offers access to the bibliographic information available on the World Wide Web (WWW). These programs are used to maintain personal reference collections and to assist with the preparation of manuscripts for publication. They are designed to be compatible with research-oriented bibliographic database services, including the library on-line public access catalogue systems (OPACs). The recent introduction of Internet Enabler expands these channels to the Internet. This product provides a link between the bibliographic information found on the WWW and ProCite. It is comprised of two modules. One module, NetCite-, accommodates the storage and cataloguing of

ADVERTISEMENT

RETROSCREEN LTD

A Virology Research & Clinical Trials Company

Specialised laboratories including category III. All work conducted to GLP-standards with external QA

Our experience includes:

- Clinical trials conducted to GLP/ICH guidelines with both vaccines and antivirals.
- *In vitro* virucidal studies of various lozenges versus flu, RSV adeno and rhinoviruses.
- R&D for virus vaccines.
- Preparation and testing of viral neutralising antibodies (HIV, influenza etc).
- Screening tests for antivirals and disinfectants against HIV, influenza and Herpes.
- Preparation of virus concentrates for HIV and influenza.

Retroscreen Ltd specialises in extensive discussion of your exact requirements, fast turnaround, precise commencement and final dates.

The London Hospital Medical College, 64 Turner Street, London E1 2AD, England.

Tel: +44 171 375 2498/375 0345

Fax: +44 171 375 2597

E-Mail: J.S. Oxford@mds.qmw.ac.uk.

www.retroscreen.com

READER ENQUIRY NO. 77

URLs within a ProCite record. The program allows the user to hyper-link from the URL stored within a record to the addressed site on the WWW. BookWhere Pro, also part of Internet Enabler, allows users to retrieve and download, into ProCite, bibliographic information from Z39.50-compliant databases accessible via the WWW.

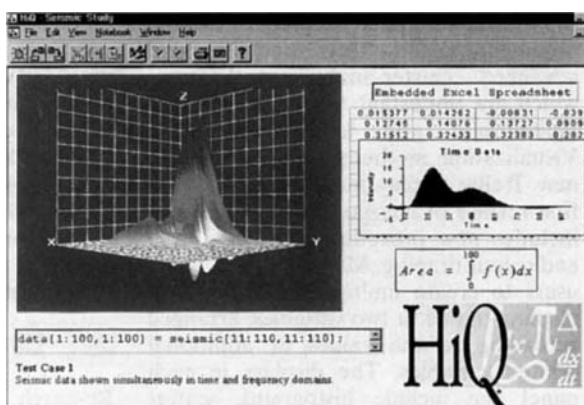
Reader Enquiry No. 114

Also in the realm of **bibliographic management**, Cherwell Scientific now offers EndNote 2.0 and EndLink 2.0, Windows versions of the original Macintosh software from Niles and Associates. These programs feature over 200 bibliographic formats for different journals; Word, WordPerfect for Windows and Ami Pro compatibility; 100 customized filters for importing references from on-line and CD-ROM services; and term-list, global editing and enhanced searching functions. EndNote 2.1 has also been released with an add-in that integrates Word for Windows (95/3.1/NT). This enables users to select a bibliography from Word's tool menu and EndNote will automatically generate a bibliography at the end of the document. Both programs are 32-bit applications that run under Windows 3.2, Window NT and Windows95. EndNote Plus 2.0 costs £299 and EndLink is £99. Upgrades from other bibliographic programs can be purchased for £99.

Reader Enquiry No. 115

Miscellaneous

National Instruments has released a multimedia **CD-ROM showcase** for Windows95/3.1 and Macintosh/Power Macintosh computers, which provides information and demonstration versions of application software packages, such as LabView, LabWindows/CVI and Measure. Also included is information on the company's data analysis and visualization tools: HiQ, Signal Processing Suite, TestSuite and LabSuite, the company's software packages for automated testing and laboratory automation applications. HiQ now offers interactive maths and data visualization capabilities, which should make advanced technical calculation and data visualization solutions more cost-effective and easier. Running on Windows95 or Windows NT-based personal computers, this 32-bit application uses the object linking and embedding (OLE) and ActiveX technologies, as well as the Microsoft OpenGLX three-dimensional graphics library to deliver its utilities

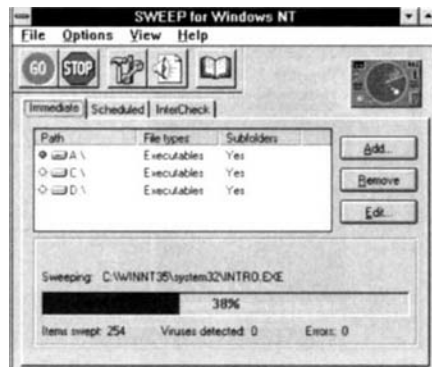


HiQ is featured in National Instruments' CD-ROM showcase for Windows 3.1, Windows95 and thePowerPC.

ties to Word, Powerpoint and Excel. Users interested in downloading a free beta version of HiQ for Windows can download it from the WWW at <http://www.natinst.com/hiq> until the shipping version is released.

Reader Enquiry No. 116

Sophos' new version of Sweep **network virus protection** runs on Windows NT for both Intel and Alpha/AXP systems. It combines virus detection technology with an intuitive user interface and full service operation to provide both convenience and security. Running as a service, Sweep for Windows NT can schedule virus checks of



Network virus protection from Sophos.

both workstations and file servers, without requiring any user to be logged in. Sweep includes InterCheck technology, which provides real-time checking of workstations. It is said to be suitable for environments where network and security managers need reliable virus detection with the minimum of user intervention. Evaluation copies of the program and technical product information can be downloaded from ([ftp.sophos.com/pub/evaluation/sweep](ftp://ftp.sophos.com/pub/evaluation/sweep)) or from the company's World-Wide Web server (www.sophos.com).

Reader Enquiry No. 117

A new optimized **Fortran 90 compiler** for Sun SPARCstations running Solaris 2 has been launched following a joint technology initiative between NAG and ACE.

The new compiler provides Fortran developers and programmers with enhanced performance and flexibility, by combining the NAGWare f90 compiler front end with three new CoSy generated back ends and optimizers from ACE. Further implementations are planned for other processors, as well as the PowerPC processors. The original NAGWare f90 compiler, achieved portability using C as its intermediate language and automatically invoked the host compiler to produce object code. The new compiler obviates the need for intermediate source-code conversion producing target-specific final code. In achieving this, the NAGWare f90 Compiler has been integrated with the CoSy parallel compilation system. The compiler also accepts HPF extensions. Back-end components in CoSy, like the code selector and code scheduler, are generated from processor descriptions, leading to rapid availability of optimized back-end systems for a variety of architectures. The generated compiler phases interact in parallel to produce optimal codegeneration strategies.

Reader Enquiry No. 118

According to Metrohm, Ti Net 2 **titration software** represents a significant improvement in their titration technology. It has been developed to leave no analytical needs unfulfilled. The system operates under Windows (3.11 or 95), and together with the Titrimo titrators, should provide users with expanded capabilities comparable with more expensive titrators. It can operate an unlimited numerous combination simultaneously of Titrimo's, showing 'live' titration curves on any four instruments. This system is designed to tackle any titration task with user-programmable parameters. It will also check system components along the way by asking important 'good laboratory practice' questions such as, 'Is the sample size correct?', 'Was the titrant calibrated within the last week?', 'Is the result on specification'. The program has a built-in database, which can be customized so that users can view the results in the format that they require. Furthermore, it allows operators to recalculate results (depending on their security access), filter out results and transfer results to other packages, such as Excel, Word, and others. Ti Net 2 is also able to operate with LIMS systems. Some other features of the software include the capability to include the use of the operator's own standard operating procedures, giving on-screen access to the protocol, the balance and bar-code reader connection.

Reader Enquiry No. 119

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.